Sílvia Alexandra Dias Almeida

Bachelor of Science

Atlas-based semi-automatic segmentation of Whole-Body Diffusion Weighted Imaging images: Quantification of tumor burden

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

Advisers: Nikolaos Papanikolaou, Principal Investigator, Champalimaud Foundation

Cristina João, Hematologist, Champalimaud Foundation, Auxiliarly Professor, NOVA Medical School

Examination Committee

Chairperson:	Prof. Dr. Carla Maria Quintão Pereira
Raporteur:	Prof. Dr. Patrícia Margarida Piedade Figueiredo
Member:	Prof. Dr. Cristina Maria Godinho Pires João

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September, 2018

Atlas-based semi-automatic segmentation of Whole-Body Diffusion Weighted Imaging images: Quantification of tumor burden

This document was created using the (pdf)LaTeX processor, based in the "novathesis" template[1], developed at the Dep. Informática of FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

To my grandparents António and Albino

Acknowledgements

It is with great pleasure and satisfaction that I finish this important step of my life. This journey would not be possible without the contribution of many people who helped me to fulfill my goals, whether professional or personal.

First, I would like to thank the Champalimaud Foundation, especially Professor Dr. António Parreira and Professor Dr. Celso Matos for having me welcomed in the institution.

To my adviser Dr. Nikolaos Papanikolaou, for the motivation, availability, confidence and for the valuable teaching transmitted during our meetings. Thank you for the opportunity to have worked on such a fascinating project.

To my adviser Professor Dr. Cristina João, for the contagious passion for science, dedication, kindness and critical mind whenever I presented you the results.

The work of scientific research is necessarily a teamwork, and, in this sense, I would like to thank Eng. João Santinha and Dr. Francisco Oliveira who worked with me in the development of this project. I fully appreciate all the support and the reception from the first day. Thank you for your critical minds, advice and for always encouraging me to explore different solutions and go beyond what was expected.

To Dr. Joana Ip, Dr. Maria Lisitskaya, Dr. João Lourenço and Dr. Aycan Uysal for your valuable segmentation work, explanations and clarifications about bone marrow involvement imaging.

To my college friends, with whom I shared moments of true happiness, joy, but also despair during these five years. The motto "Tudo se faz" has really brought us here, right?

To my partner, thank you for all the love, but most importantly, the friendship. Thank you for the long conversations, for sharing our fears and accomplishments, for the support, stability and for growing side by side.

Last but not least, I would like to thank my family, especially my parents and sister, for the everyday support, care and encouragement. Thank you for being my inspiration and for value the importance of this document.

Abstract

Cancer is a leading cause of death worldwide. Treatment strategies rely on accurate tumor staging and surveillance by imaging screening. Whole-body Diffusion Weighted Imaging (DWI) has high value to detect, characterize and quantify malignancies with irregular diffusion patterns, such as Multiple Myeloma (MM). However, the large volume of imaging data hinders the reading process. Manual delineation (segmentation) of tumor sites becomes a time-consuming process and lacks reproducibility. The lack of adequate tools in clinical practice leads radiologists to perform only a qualitative description of DWI images and measure of the biggest lesion diameter, an inherently subjective process. Arising from this need, this dissertation aimed to develop an algorithm to improve the process of segmentation of lesions of MM DWI images, to allow accurately and rapidly tumor burden quantification, by validating against radiologist's manual segmentation. Quantification of bone lesions (hyperintense on DWI) volume without considering normally hyperintense organs was made possible due to the development of an atlas-based and a smart lesion detector algorithm. The first allowed the removal of normal hyperintense organs from the images to be studied, using a suitable registration procedure. The second applied an outlier detector algorithm and compared voxel-by-voxel and connected-component approaches on different b-value images (directly acquired and computed), to delineate lesions. T1-weighted images were also used to improve lesion detection. The atlas-based algorithm revealed good alignments against the manual segmentation: Dice Similarity Coefficient (DSC) of 0.63±0.03 for male and 0.58±0.05 for female. Regarding lesion detection, the connected-component approach applied to the directly acquired b-value image was the method that presented the greatest similarity to the gold standard. Although not yet overcoming the manual segmentation performance, these results are suggestive of the great potential of semi-automatic registration methods combined with quantitative algorithms to analyze DWI images, assisting radiologists while defining tumor burden. Staging, prognosis and response analysis in several pathologies may be facilitated.

Keywords: Diffusion Weighted Imaging, Semi-Automatic Segmentation, Atlas-based Segmentation, Tumor Burden, Image Processing

Resumo

O cancro é uma das principais causas de morte no mundo. As estratégias de tratamento dependem do estadiamento preciso e da vigilância por exames de imagem. A Imagem Ponderada em Difusão de corpo inteiro (DWI) tem valor elevado para detetar, caracterizar e quantificar doenças com padrões de difusão irregulares, como o Mieloma Múltiplo (MM). No entanto, o volume elevado de imagens dificulta o processo de leitura. O delineamento manual (segmentação) dos tumores torna-se um processo demorado e carece de reprodutibilidade. A falta de ferramentas adequadas na prática clínica leva os radiologistas a descreverem apenas qualitativamente as imagens de DWI e a medir o maior diâmetro da lesão, um processo inerentemente subjetivo. Decorrente desta necessidade, a presente dissertação teve como objetivo desenvolver um algoritmo para melhorar o processo de segmentação de lesões em imagens de DWI de doentes com MM, permitindo a quantificação rápida e precisa da carga tumoral, comparando com a segmentação manual de radiologistas. A quantificação correta do volume de lesões ósseas, sem considerar órgãos normalmente hiperintensos, foi possível devido ao desenvolvimento de um algoritmo baseado num atlas e um detetor inteligente de lesões. A primeira permitiu a remoção dos órgãos hiperintensos normais das imagens a serem estudadas, utilizando um alinhamento adequado. O segundo aplicou um algoritmo de deteção *outlier* e comparou as abordagens voxel-por-voxel e componentes-conectados em diferentes *b-values*. O algoritmo baseado no atlas revelou bons alinhamentos comparando com a segmentação manual: coeficiente de similaridade de Dice (DSC) de 0.63±0.03 para homens e 0.58±0.05 para mulheres. Em relação à deteção de lesões, a abordagem de componentes-conectados aplicada à imagem de *b-value* diretamente adquirida foi o método que apresentou maior similaridade com os radiologistas. Apesar de ainda não superar o desempenho da segmentação manual, os resultados sugerem o potencial dos métodos de alinhamento semiautomáticos, combinados com algoritmos quantitativos para analisar imagens DWI. O estadiamento, prognóstico e análise de resposta poderão ser facilitados em várias patologias.

Palavras-chave: Imagem Ponderada em Difusão, Segmentação Semiautomática, Segmentação baseada num Atlas, Carga Tumoral, Processamento de Imagem

Contents

List of Figures

List of Tables

Acronyms

AB	Atlas-based.
ADC	Apparent Diffusion Coefficient.
AHD	Average Hausdorff Distance.
CCC	Champalimaud Clinical Centre.
cDWI	computed Diffusion Weighted Imaging.
CT	Computed Tomography.
DSC	Dice Similarity Coefficient.
DWI	Diffusion Weighted Imaging.
F18-FDG PET	F18-Fluorodeoxyglucose Positron Emission Tomography.
FID	Free Induction Decay.
HD	Hausdorff Distance.
IMWG	International Myeloma Working Group.
LoG	Laplacian of Gaussian.
MGUS	Monoclonal Gammopathy of Unknown Significance.
MI	Mutual Information.
MM	Multiple Myeloma.

MR	Magnetic Resonance.
MRI	Magnetic Resonance Imaging.
NCC	Normalized Cross Correlation.
NIfTI	Neuroimaging Informatics Technology Initiative.
NPV	Negative Predictive Value.
PD	Proton Density.
PET	Positron Emission Tomography.
PGSE	Pulsed Gradient Spin Echo.
PPV	Positive Predictive Value.
RF	Radiofrequency.
SA	Smart Algorithm.
SMM	Smoldering Multiple Myeloma.
SNR	Signal to Noise.
SPECT	Single-Photon Emission Computed Tomography.
SS-EPI	Single Shot Echo Planar Imaging.
SSD	Sum of Squared Differences.
STIR	Short Tau Inversion Recovery.
TE	Echo Time.

TR	Repetition Time.
WB	Whole-Body.
WBLD-CT	Whole Body Low Dose Computed Tomography.
WBXR	Whole Body X-ray.

Chapter 1

Introduction

1.1 Context and Motivation

Accurate tumor staging and surveillance are critical when it comes to design optimal treatment strategies of malignant diseases. After confirming the diagnosis of a neoplastic disease, precise tumor localization and description of the degrees of organ infiltration are of paramount importance. Moreover, in order to assess the prognosis and to evaluate the response to therapy in a patient with a neoplastic disease, it is crucial to perform a precise identification and characterization of malignant lesions.

The anatomy of the human body has been subject of many studies and can be mapped in a non-invasive way by Magnetic Resonance Imaging (MRI), Computed Tomography (CT), digital mammography and other imaging modalities, facilitating diagnosis, staging of malignancies and treatment planning. In particular, MRI provides anatomical images of the human body with a high spatial resolution and superb soft tissue contrast while there is no ionizing radiation exposure for the patient. Recent technological advancements in MRI made feasible the fact that Whole-Body (WB) MRI can be performed in the clinic in a reasonable examination times and without compromised image quality. Among these technological advancements are: utilization of more powerful and faster gradients, advances in hardware, use of a rolling platform-moving technology, phased array coils and parallel acquisition techniques [1].

Diffusion Weighted Imaging (DWI) is a novel and promising MRI technique based on the principle of diffusion observed in water molecules. By measuring the total lesion volume on DWI images, disease burden can be quantified. Combining DWI to WB protocols, allows the use of WB-MRI modality for treatment response assessment, besides cancer diagnosis and staging. Nevertheless, this modality results in an increased volume of imaging data, due to the reading of the WB. Consequently, even for a trained

expert, the reading process and manual delineation of the regions of interest becomes a painstakingly slow process, prone to error, hard to reproduce, time consuming with an increased risk of misinterpretation. Therefore, manual segmentation of large data sets is not a feasible solution [2]. Thence, nowadays, the gold standard for interpretation of DWI is a qualitative description by an expert radiologist.

Fully-automatic and Semi-Automatic Segmentation algorithms have been developed over the years, in order to address this problem, speeding up and removing bias from the process. These approaches are decisive in terms of quantification and localization of lesion, diagnosis, study of the anatomical structure, treatment planning and computer--integrated surgery [3]. Semi-automatic methods are limited since they require several user's interventions, especially in the initialization step. Automatic identification and detection of structures in images using registration-based segmentation have been accepted as a viable approach. The process involves a spatially normalized fitting of each image under study to the template image (or vice versa), facilitating posterior analysis [4]. Ideally, the template image is an average of the images under study, resulting in a better representation of the data available, and thus a better registration. Using just one image as the template would result in uncertainties and displacements in the registration. Atlas-based (AB) segmentation approaches have also been used to fully automatic segmentation, which exploits already segmented template images (atlas images).

Despite the large number of studies about image segmentation and registration, these methods still need improvement regarding the difficulty of quantitative validation and adaptation to large data sets. DWI, in particular, has recently started to be studied for lesion detection using semi- or automatic segmentation and registration methods [2, 5, 6, 7]. However, some of these studies have not been validated against expert manual segmentation. Therefore, there is a need for developing, improving and integrating novel registration and segmentation methods in DWI, in order to delineate specific regions pre- and post-treatment, assessing the response to therapy in neoplastic diseases.

1.2 Objectives and Dissertation Plan

The primary objective of the dissertation is the creation of a novel AB semi-automatic segmentation method for removal of hyperintense organs in WB-DWI images. The secondary objective is to develop a smart segmentation algorithm of bone lesions in WB-DWI and to validate the algorithm in a cohort population of a group of neoplastic patients, diagnosed with Multiple Myeloma (MM), in different stages of the disease. It is also intended to validate by comparing the smart semi-automatic lesion detection to the expert's manual segmentation. The main goal of the project, of which this dissertation is part, is to develop a smart segmentation tool (figure 1.1) that, in the future, could be used by radiologists to assess tumor burden.

To do so, the present dissertation has been structured so as to first cover the current state-of-the-art concerning medical image registration and segmentation, and recent

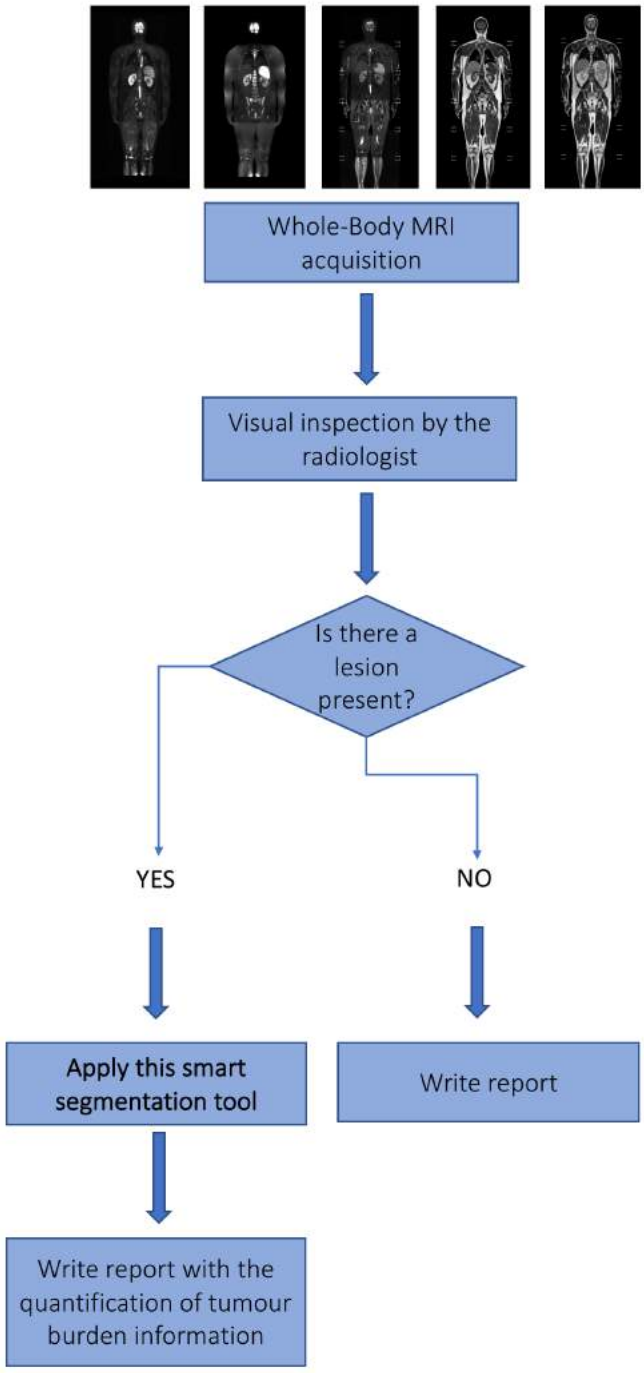


Figure 1.1: Flowchart of a future MRI analysis, by the radiologist. By introducing a smart segmentation tool, that helps to delineate lesions in semi-automatic way, this information can complement the written report. After a visual inspection of the MRI images by the radiologist, if a suspected lesion is found, confirmation and quantification may be given by applying the smart segmentation algorithm, developed in this thesis. If the radiologist does not find any suspected lesion, he writes a report stating so. Otherwise, tumor burden information is presented in the report.

methods of accurately delineate tumor volume and further assessment of treatment response in WB-DWI. In chapter 2 are described some of the most common segmentation and registration techniques currently in use, and their specific advantages and disadvantages are highlighted. Afterwards, the relevant theoretical underpinnings of MRI and diffusion are explained. In chapter 4, the reader will be presented with an explanation about MM: symptoms, cause, resumed pathophysiology, and diagnosis. In the following chapter, the materials and methods used for imaging processing are presented. Then, the results and discussion will be presented. Lastly, are described the conclusions, limitations and future perspectives, which summarize the dissertation's findings and suggest the next steps of this study.

1.3 State-of-the-Art

Medical image analysis emerged in the early 1990s, as a branch of Artificial Intelligence and Computer Science. In the beginning, mathematical methods that had gain traction in non-medical imaging problems, were applied to medical images [8]. In the following years, since the development of imaging techniques and digital imaging revolution, there was a need to enhance and delineate specific regions of the human body, in order to compare with other images or to quantify size or volume or to better study structures. Consequently, studies encompassing segmentation algorithms were rapidly developed.

Segmentation is one of the major problems in medical image analysis and consists in the process of subdividing a digital image into sets of pixels or voxels, tagging them with biological meaningful labels. For most applications, it remains as a manual task, where the expert sketches the contours slice by slice, using pointing devices like a mouse or a trackball. Thus, this approach is time consuming and prone to inaccuracies introduced by the user. Accuracy and precision are of great relevance to medical image segmentation, particularly to assess a prognosis and evaluate the response to therapy in neoplastic diseases. Therefore, specific segmentation algorithms are required in order to fulfill this gap in medical image analysis. Segmentation algorithms may differ from imaging modalities and different slices, since the appearances of a given organ may vary. Consequently, when designing an effective algorithm, it must be taken into account the imaging modality, structures to analyze, influence of noise and partial volume effects [9].

In addition to segmentation, registration is also of paramount importance in medical image analysis. It consists on the alignment of two or more images, allowing matching and comparison, fusion of different imaging modalities (e.g. MRI and CT) or to highlight structures of interest to facilitate further segmentation. The alignment is based on a transformation model, which defines the geometric transformation applied between images.

Over the years, several algorithms for medical image segmentation and registration were developed, answering to different problems in medicine. The majority of articles focus on brain [10, 11, 12, 13], but in general, using different imaging modalities, all body

parts are object of studies: bones [14, 15, 16], heart [17], abdominal organs [18, 19] and kidneys [20].

AB segmentation has become one of the most prominent approaches to semi- and fully-automatic segmentation. An atlas, in medical imaging, is a template image that represents all the images available for study. Usually, it is based on pairwise registration, where an image is selected as a reference and the other images are registered to it, one at a time, followed by an average of the result. Once the atlas is created, one can access the average location of human structures, such as bones, organs or tissues. Knowing the structures' average position, segmentation can be done, either manual or automatic. Then, the segmented label can be transferred to the same physical space as a new image to be segmented, by applying the same transformation model used to register the images to the reference.

The usefulness of AB segmentation has been shown in many studies [21, 22] and it is not restrained to an imaging modality. The introduction of multiple images as atlas - multi-atlas segmentation - can improve the representation of inter-subject variability, compared to the use of a single atlas coupled with a deformation model [23].

Regarding WB-MRI, there have been developed some methods to fully automatic segment regions of the body. In order to assess measurements of the muscle volume, A. Karlsson *et al* [24] developed an automatic multi-atlas method for the quantification of total and regional skeletal muscle volume. This process involved segmentation of intensity-corrected water-fat separated image volumes, non-rigid registration of multiple atlas, muscle tissue classification and volume quantification. The validation showed high accuracy compared to manual segmentation. I. Lavdas *et al* [2] developed and evaluated three automatic methods of segmentation of organs and bones in WB-MRI, including a multi-atlas approach. This approach included an intensity-based image registration, with free-form deformations as the transformation model and correlation coefficient as the similarity measure. Although it did not improve the segmentation performance, compared to the gold standard manual segmentation, it is the foundation for the development of robust algorithms for the automatic detection and segmentation of lesions in WB-MRI scans.

MM is a hematologic neoplasia characterized by clonal abnormal plasma cells in the bone marrow and/or in extramedullary sites, leading to hypercalcemia, renal insufficiency, anemia and osteolytic lesions. Usually, tumor deposits occur on appendicular skeleton (proximal and long bones) but can also occur on the axial skeleton. The International Myeloma Working Group (IMWG) holds that MRI is the gold standard for detection of bone marrow involvement in MM [25]. Normal, focal, diffuse, a combination of both focal and diffuse, and variegated are the typical patterns found on MRI of MM patients, when there is marrow involvement [26]. Diffuse pattern is often associated with advanced disease and worse prognosis [27, 28].

Apparent Diffusion Coefficient (ADC) is a diffusion extracted imaging biomarker that reflects the magnitude of diffusivity of water molecules within biological tissues.

Changes in ADC value can be related to treatment response. DWI is commonly applied in oncologic imaging, helping to detect and characterize malignancies that show irregular diffusion behaviors. MM lesion's patterns, for instance, have been recently evaluated by ADC to help distinguish a diffuse from a normal pattern [29]. The results showed that ADCs greater than 0.548×10^{-3} mm^2/s indicated 100% sensitivity and 98% specificity for the diagnosis of a diffuse pattern. Before extracting ADC metrics of total or partial lesion volume, tumor volume must be extracted in DWI. Traditional methods include manual delineation of the lesion, slice by slice. Yet, this is not applied in the common clinical practice: usually, the radiologist selects the larger sized lesion, measures its biggest diameter and extracts its ADC. This process can be considered very subjective when follow-up is performed: even if performed by the same radiologist, the diameter measured might not be the same. Plus, lesions are likely to not have a sphere-shape, which will translate into differences in diameter measures due varying positioning of the patient, between baseline and follow-up.

Adding to the fact that manual segmentation of lesions is a very difficult process, lacking high grades of agreement, it is estimated that radiologists have a 3-5% real-time day-to-day error rate and the retrospective error rate among radiologic studies is 30% [30]. To our knowledge, there is no study that compares the disagreement between lesion delineation in DWI-MRI by different radiologists, with different levels of experience.

Studies encompassing accurate tumor volume automatic delineation and further assessment of treatment response in WB-MRI are very few. Blackledge MD *et al* [7] developed a semi-automatic segmentation algorithm of WB-DWI using a Markov random field model to infer tumor volume and associated global ADC. However, it still requires a lot of user interventions to define contrast between lesion and normal tissues and to define a threshold that covers the lesions. Also, it lacks on validating the segmentation algorithm, since they only evaluated based on responder/non-responder to treatment. Plus, the associated computational time is of the order of 30 minutes, which is considered long.

Therefore, given that the expert manual segmentation of lesions is a very difficult process to reproduce and given the small number of studies regarding semi-automatic tumor volume quantification in DWI and corresponding response to treatment, it is important to fill this gap in research. The development of an algorithm that has short computational time, requires few user interventions and effectively facilitates the staging process, prognosis and analysis of response, would have direct clinical implications.

1.4 Dissertation Outputs

This dissertation has developed some output that was accepted for presentation in scientific events, which are worth mentioning:

Communication

- Almeida, D. S., Santinha, J., Oliveira, F., Papanikolaou, N., João, C., "Atlas-based semi-automatic segmentation of Diffusion Weighted Imaging", 3rd NOVA Biomedical Engineering Workshop (NBEW), 9th May 2018, Caparica, Portugal.

An article is also being prepared as an output of this thesis, to be submitted during this year.

Chapter 2

Image Registration and Segmentation

2.1 Image Registration

Image Registration, also known as image fusion, warping or matching, is a computationally expensive task that involves deforming (using a suitable deformation model) an image until it is similar to another image. The purpose of this method is to find the optimal transformation that produces the best alignment of the structures of interest in the input images. The input images comprise a reference (fixed) and a moving image, that will be aligned with the reference one. Medical imaging, target recognition, cartography and computer vision are some of the main applications of this method [31].

Registration is a crucial step for medical image analysis since it is necessary in order to compare images from different sensors or modalities (multi-modal image registration), different viewpoints or even acquired at different times (serial image registration). Registration can be applied intra-subject or inter-subject, when, e.g., the goal is to compare a certain characteristic in a given population.

Over the past decades, image registration has been the scope of several studies in medical image analysis [32], regarding fusion from anatomical images from CT, fusion of functional images from MRI and Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT) or functional MRI. It can also be applied to diagnosis, staging, assessment of treatment response, detection of disease recurrence, surgery simulation, generation and comparison of atlas, radiotherapy, anatomic segmentation, comparison of images pre- and post-contrast injection and many others.

Typically, the process of image registration (figure 2.1) involves the following components:

- A *transformation model*, responsible for defining the geometric transformation applied between images. Usually, are divided into non-deformable and deformable

classes. The choice of the geometric transformation depends on the nature of the images to be registered.

- A *similarity measure* or alignment quality, which measures the degree of alignment between images. The methods for the alignment measure depend on the information contained in the image. Thus, measures can be obtained in terms of intensity - intensity-based similarity measures - or in terms of shapes, edges, landmarks - feature-based similarity measures.

- An *optimization method*, which searches for the maximum or minimal value of the similarity measure adopted. The goal is to find a transformation that correctly registers the input images.

- An *accuracy assessment* protocol, which measures the performance of the registration protocol in terms of accuracy and reliability.

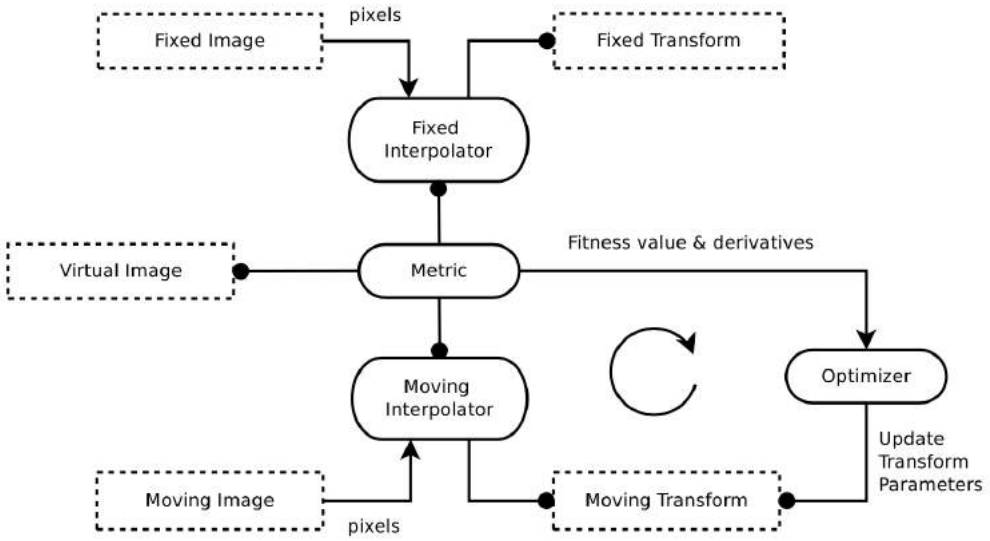


Figure 2.1: Typical registration methodology. The main idea is to search iteratively for the transformation model that optimizes the similarity measure, when applied to the moving image. The interpolator resamples the voxels/pixels of the moving image into the new coordinate system, defined by the search strategy (metric) found by the optimizer. Adapted from [33].

Usually, a simple pre-registration method is applied before the registration step, in order to obtain an initial solution. The moving image becomes closer to the fixed image in terms of the similarity measure, which allows a faster convergence of the optimizer and decreases the chance of an erroneous solution. The transformation model, similarity measure, optimization method and accuracy assessment protocol will be discussed in this chapter, as well as some of the applications in the clinic. Also, throughout this chapter, $\mathbf{X}$, $\mathbf{x}$ and x will be used to indicate a matrix, vector, or scalar quantity, respectively.

The following content was mostly based on the literature review published by Oliveira & Tavares [4], B. Zitová *et al* [34] and Rueckert & Schnabel [35].

2.1.1 Transformation Model

The transformation model is given by equation 2.1,

$$\mathbf{T} : (x, y, z) \rightarrow (x', y', z') \tag{2.1}$$

whose goal is to align each point in the moving image with the reference one.

Geometric transformations, defined by the transformation model, can be classified as non-deformable and deformable. Among non-deformable transformations, there is translation, rigid and affine transformations. Rigid transformation can be used when correspondence is achieved by simply rotation or translation of an image. It is defined in the 3D space by six degrees of freedom. Deformable transformations are applicable when correspondence between structures can only be achieved by stretching the image or other more complex transformation. Similarity (translation, rotation and uniform scaling), affine (translation, rotation, scaling and shear), projective and curved transformations are included in the deformable transformation class.

The transformation model chosen must be able to characterize the geometric transformation between images, i.e., represent the transformations involved between them. For example, when registering rigid structures, as bones, of the same subject, there must be employed a non-deformable transformation. On the other hand, images that are affected, e.g., by respiratory motion must not be registered with a non-deformable transformation, since it does not represent the transformations required to align and deform the images. Plus, the model must be as simple as possible.

Both rigid and affine transformations can be characterized by a small number of parameters, so they are considered simple. Rigid transformations are defined by three translational and three rotational parameters and can be applied in two situations. The first is for the registration of rigid structures, such as bones [36], [37]. The second is for pre-registration before a more complex transformation [38]. Affine transformations are defined by twelve parameters, represented by rotation, translation, scaling and shears. Thus, affine registration is more complete than rigid, since it enables corrections for scaling and shears. Rigid and affine transformations are suitable for registration of anatomical structures, like the brain and bones.

Since most anatomic parts are deformable structures, rigid and affine transformations tend to be inflexible. Soft tissues' images registration problem should be addressed with robust deformable transformations, also known as non-rigid or curved transformations. There are two kinds of deformable transformations: free-form and guided. Free-form transformations consist on the definition of a grid of controlling points, which are moved individually in a direction that optimizes the similarity measure. The deformation between control points can be propagated by linear interpolation or other convex kernels.

Usually, it is interpolated by the cubic B-spline, which allows very good alignments with high computational efficiency for a larger number of control points, since they are locally controlled. This means that when a control point is moved to a new position, only the point of the new position is affected, unlike global solutions (e.g. Thin-plate splines). Even though being locally controlled, B-splines can also be classified between a global registration and a local model, since they are controlled by a varying grid of controlling points. Thus, it is important to correct the global misregistrations before using free-form B-splines, for instance, with an affine transformation.

In guided transformations, the deformation is controlled by a physical model that takes into account the material properties, as the elasticity and fluid flow. This physical model treats the source image as a linear, elastic solid and is ruled by internal and external forces. Internal force is opposed to deformation of the material from its equilibrium shape, whereas external force acts on the moving image. Therefore, the deformation of the moving image stops when both forces reach an equilibrium. Local correlation measure based on intensities, intensity differences or any gradient of a similarity measure mentioned in 2.1.2 can be used as an external force. Though, the linear elasticity assumption is only valid for small deformations, so the elastic model is usually replaced by a viscous fluid model, also known as fluid-based algorithm. In this model, the registration problem is addressed as a motion problem, i.e. the content of an image is moved continually into the other, driven by the minimization of energy of the physical model adopted.

Finally, in order to preserve the topology of the structures represented in the images to be registered, the geometric transformation must be diffeomorphic. This means that it must be invertible and maps one differentiable manifold to another, so that both the function and its inverse are smooth.

2.1.2 Similarity Measure

After the selection of the geometric transformation between the images, the alignment between them must be measured. This measure of similarity is divided in two classes: the intensity- and the feature-based methods.

Intensity-based methods, also known as voxel-based methods, aims at measuring the degree of shared information between the images' intensities. The most commonly used are based on intensity differences, intensity cross-correlation and information theory.

Intensity differences measurements are based on the Sum of Squared Differences (SSD) (equation 2.2), or their normalization, between intensities in the images Γ_A and Γ_B,

$$S_{SSD} = \frac{1}{n} \sum ((\Gamma_A(\mathbf{q}) - \Gamma_B(\mathbf{T}(\mathbf{p})))^2 \tag{2.2}$$

where $\mathbf{p}$ and $\mathbf{q}$ are two-point sets, S_{SSD} is the similarity measure based on SSD, and n is the number of voxels in the region of overlap. This method is based on the assumption that the corresponding structures in both images have identical intensities, so it is restricted

to mono-modal applications. Ideally, if the images are correctly registered and, therefore, well aligned, S_{SSD} is zero.

Intensity cross-correlation is a more general approach, based on the assumption that there is a linear relationship between the intensities of the images to be registered. It can be applied to some multi-modal tasks, but the majority is applied to mono-modal. The Normalized Cross Correlation (NCC) is defined as follows in equation 2.3,

$$S_{NCC} = \frac{\sum(\Gamma_A(\mathbf{q}) - \mu_A)(\Gamma_B(\mathbf{T}(\mathbf{p})) - \mu_B)}{\sqrt{(\sum \Gamma_A(\mathbf{q}) - \mu_A)^2 (\sum \Gamma_B(\mathbf{T}(\mathbf{p})) - \mu_B)^2}} \tag{2.3}$$

where μ_A and μ_B are the voxel's intensities average in the images Γ_A and Γ_B, respectively. The larger the S_{NCC} is, the better registered the image is.

In information theory, the images'content can be described as the Shannon-Wiener entropy, $H(\Gamma_A)$ and $H(\Gamma_B)$ of images Γ_A and Γ_B, computed from the joint probability distribution of the image voxel intensity. Equations 2.4 and 2.5 describe Shannon-Wiener entropy for both images,

$$H(\Gamma_A) = -\sum_{a \in \Gamma_A} p(a) \log p(a) \tag{2.4}$$

$$H(\Gamma_B) = -\sum_{b \in \Gamma_B} p(b) \log p(b) \tag{2.5}$$

where $p(a)$ and $p(b)$ are the probabilities that a voxel in images Γ_A and Γ_B has intensity a and b, respectively. Measurements of alignment can be obtained by the information content of the joint histogram obtained from the fixed and moving image, i.e., entropy of the joint histogram. The feature space of the image intensities can be seen as the joint probability distribution. The joint Shannon-Wiener entropy of the joint probability histogram $H(\Gamma_A, \Gamma_B)$, of images Γ_A and Γ_B, may be defined by equation 2.6,

$$H(\Gamma_A, \Gamma_B) = -\sum_{a \in \Gamma_A} \sum_{b \in \Gamma_B} p(a,b) \log p(a,b) \tag{2.6}$$

where $p(a,b)$ represents the joint probability that a voxel in the overlapping region of the image Γ_A and Γ_B has a and b as values, respectively. Information theory is mostly based on Mutual Information (MI), which can be defined as follows in equation 2.7.

$$S_{MI}(\Gamma_A; \Gamma_B) = H(\Gamma_A) + H(\Gamma_B) - H(\Gamma_A, \Gamma_B) \tag{2.7}$$

MI is a measure of how well does one image explains the other image, so it is based on the supposition that there is a functional between the variables involved, e.g. the intensities. Therefore, MI is maximum when the images are correctly aligned. MI fails to consider relevant spatial information intrinsic to the original images since the computation is voxel by voxel, so only the relationships between corresponding individual voxels are considered.

Feature-based methods usually establish points, lines or surfaces in the two images in order to minimize the distance between corresponding features. The similarity measure

is often computed using the sum of the distances between two fixed points, $\mathbf{p}$ and $\mathbf{q}$ (equation 2.8).

$$S = -\sum_{i} \| \mathbf{q}_i - \mathbf{f}(\mathbf{p}_i) \|^2 \tag{2.8}$$

Since points tend to be relatively sparse, surfaces of anatomical structures are commonly used when more dense features are required. Segmentation or landmark detection of the contours structures is a primary step in order to extract the features of the images. The resulting contours are represented as point sets, which can be registered by minimizing the distance between corresponding points of both sets. The correspondence between point sets needs to be known beforehand. One of the advantages of this method is that it can be applied to mono- and multi-modality registration. Since there is a need for feature extraction, some bias could be introduced if it is done manually, propagating the error to the registration process. As seen before, intensity-based methods have advantages over feature-based in this matter since does not require any feature extraction.

2.1.3 Optimization Method

Image registration problem can be assumed as an optimization problem, whose goal is to optimize an objective function. Frequently, the objective function is composed by two terms: the similarity measure between the images and a penalty term due to the geometric transformation. In the case of rigid or affine registration, the optimization algorithm is simply maximizing the similarity metric, since the last term plays no role. However, an affine registration can introduce unacceptable deformations. For non-rigid registration, the second term has a prominent role since it represents a prior knowledge about the expected transformation.

2.1.4 Accuracy Assessment

Validation of the registration algorithms are of great value in medical image analysis. Since the optimization problem relies on multiple adjustable parameters, the accuracy also depends on that choice.

There are several approaches to measure the accuracy of registration algorithms. As a first approach, the image similarity optimization could be used as a simple accuracy measure since in image registration the problem is defined as an optimization problem. Yet, it has little value in terms of geometry, so it is rarely used.

A second approach is to apply a transformation model to an image and then use the registration algorithm to realign both images. Knowing the transformation model applied, one can infer about the accuracy of the registration process. In a similar way, by simulating the imaging acquisition physics which results in a synthetic image, and then evaluate the registration algorithm in the image produced. Manual identification of fiducial markers for registration accuracy are also reliable solutions.

Points correspondence between images are an important measure of accuracy. Euclidean distance gives a physical value of the relation between correspondence points, so it is commonly used to assess accuracy.

2.2 Image Segmentation

Image segmentation is the process of extracting meaningful information from an image, separating it into several components. This method can be applied through an automatic or semi-automatic process. The difference between them is that semi-automatic segmentation requires user initialization and correction. Many image segmentation methods have been applied to medical image analysis, facilitating the visualization and border detection of tissues and body organs.

Several image segmentation techniques exist, which can be divided in algorithms based on thresholding, edge-based, region-based, clustering, deformable models and AB. Similar to registration, segmentation methods also need to evaluate the performance of the algorithms, and so to assess the accuracy.

For instance, Dice Similarity Coefficient (DSC) is commonly used to evaluate the agreement between binary images, quantifying the matching between overlapping regions. This measure is frequently used to assess the degree of overlap of two segmentation. Equation 2.9 defines the Dice's formula, where A and B are two datasets. DSC equals twice the number of common elements of both groups divided by the sum of elements in each group. Thus, DSC=1 means that there is a complete overlap, while DSC=0 means that there is not one single element in common between the datasets.

$$DSC = \frac{2*|A \cap B|}{|A+B|} \tag{2.9}$$

The following content was mostly based on the literature review published by L. K. Lee *et al* [39] and E. Neri *et al* [40].

2.2.1 Thresholding-based

Threshold-based segmentation methods are the simplest and straightforward methods, which are based on the assumption that images are formed from regions with different intensities. By analyzing the histogram of each images, if the intensity value is greater than some threshold, the corresponding pixels are targeted (foreground). If the value is lower than the threshold, the corresponding pixels are considered background. Equation 2.10 represents the threshold-based method,

$$g(x,y) = \begin{cases} foreground & f(x,y) \geq T \\ background & f(x,y) < T \end{cases} \tag{2.10}$$

where $f(x,y)$ is the pixel intensity in the (x,y) position and T is the defined threshold value. In order to segment an image with different intensities, more than one threshold

value can be applied. This approach is considered global, since is based on the assumption that an object can be separated from the background using a threshold value. However, it fails when an image does not have a constant background, and only a region can be successfully segmented. In order to solve this problem, there can be applied local approaches. These approaches divide images into sub images and calculate the threshold for each one, and then the results are merged. Mean and standard deviation can be used to select the threshold value for each sub-image.

More sophisticated techniques have been developed, such as the Otsu's method [41], which makes an initial guess of the thresholds and then maximizes the separation between different threshold classes in the data.

Thresholding can be combined with a connected-component analysis, which scans a binary image pixel-by-pixel from top to bottom, left to right, and identifies connected pixels, until a tissue type has been labelled. Then, a new threshold value is applied and the process is repeated for another tissue type.

2.2.1.1 Outlier Removal

Instead of defining a fixed numerical value for all images, above which segmentation occurs and the fact that some of the MRI sequences produce non-quantitative images, the value can be defined based on each image's intensity distribution. For instance, when searching for specific regions in an image that stand out in terms of intensity (hypo or hyperintense), the image's histogram can be calculated and the extremes (outliers) can be extracted.

The Tukey's fences [42] is one of the several methods to detect outliers [43, 44, 45]. If $Q1$ and $Q3$ are the lower and upper quartiles, respectively, than the upper outlier range can be defined as a constant (k) multiplied by the interquartile range (difference between $Q3$ and $Q1$) plus $Q3$, and the lower outlier range as the difference between $Q1$ and k multiplied by the interquartile range. k is a non-negative constant. Tukey *et al* suggest using $k = 1.5$ to define an outlier and $k = 3$ to define a range above that. However, when applied to imaging processing, this constant should be manipulated according to the size of the image and to highlight/remove.

Equations 2.11 and 2.12 define the outlier equations and figure 2.2 shows a distribution divided by quartiles.

$$\text{Upper outlier range} \geq Q3 + k(Q3 - Q1) \tag{2.11}$$

$$\text{Lower outlier range} \geq Q1 - k(Q3 - Q1) \tag{2.12}$$

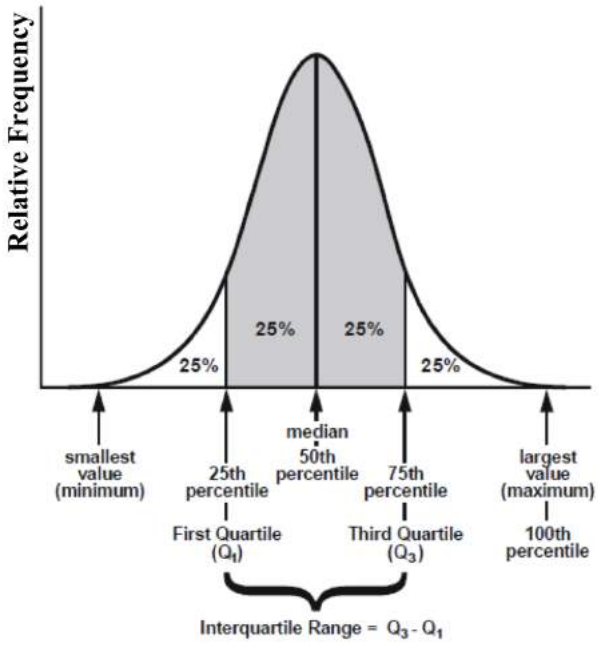


Figure 2.2: Representation of a distribution divided by quartiles. The first 25% is the first quartile ($Q1$), followed by the second quartile that represents the median ($Q2$) and the third quartile ($Q3$), corresponding to the 75% percentile. The interquartile range is defined as the distance between the third and first quartile. Adapted from [46]

2.2.2 Edge-based

Edge-based segmentation methods are based on the intensity variations presented in the borders of regions in an image. Sobel filters are gradient operators commonly applied to identify and extract borders.

The magnitude and orientation of an edge can be estimated by means of the Prewitt edge detector [47]. This operator calculates the gradient of the image intensity at each point, resulting in the direction of the largest possible increase from bright to darker values and the rate of change in that direction. Thus, the Prewitt edge detector provides information about how abruptly the image changes in a particular point, and thus how likely it could represent an edge, as well as its orientation.

The Laplacian is a 2-D isotropic measure of the second spatial derivative of an image. Applied to an image, it highlights regions of abrupt intensity change and so it is often used for edge detection. Gaussian filters combined with the Laplacian technique are also segmentation edge-based detectors, often referred as Laplacian of Gaussian (LoG). This method results in a reduced sensitivity to noise, since the image has first been smoothed by the Gaussian filter. Since the LoG operator calculates the second intensity derivative of an image, the areas with constant intensity correspond to LoG equal to zero. On the other hand, when there is a change in intensity of the surroundings of a region, LoG will be negative on the lighter side and positive on the darker side.

Another example of edge-based methods are watershed algorithms. By combining image intensity with gradient information and using mathematical morphology operations, they can divide images into homogenous regions. These homogenous regions are the pixels enclosed in the same watershed line, which are defined by the pixels with local maximum gradient magnitudes. Watersheds produce efficient segmentation due to the incorporation of diverse image information. However, these algorithms tend to suffer from over-segmentation, especially when the images are noisy or the desired objects have low signal-to-noise ratio appearances [9]. In medical image segmentation, this method is usually followed by a post-processing step to merge regions that were separated but belong to the same structure.

2.2.3 Region-based

Region-based segmentation algorithms consider the image to be analyzed as a homogenous region and aim to search for the pixels with similar feature values. Region Splitting and Merging and Region Growing are examples of region-based segmentation. The first, considers the image as a single region and then determines if the homogeneity criteria are satisfied. If not, the region is divided into smaller sub-regions. The process finishes when no further division is required. Then, the resulting sub-regions are compared and merged if they are similar.

Region growing algorithms incorporate the use of seed points, manually identified in the images. Then, a homogenous region is grown around the fixed seed points and neighboring pixels with similar intensities are iteractively added. The criteria adopted to decide if a pixel should or not be added and the connectivity used to determine neighbors depends on the algorithm adopted. The criteria behind the homogenous region growth can be threshold-based, i.e., select all the neighbors with intensity lower than or equal to the seed point's intensity and above a threshold. The suitable threshold value can be obtained by a trial-and-error process or by the analysis of the image's histogram.

2.2.4 Clustering-based

Clustering methods are unsupervised pattern recognition techniques that aim to divide and segment an image, without the use of training data. For this reason, clustering methods iteractively alternate between segmentation and characterization of the regions. Therefore, using only the data available, these methods can train themselves [3]. In this section, the most popular clustering algorithms will be explained, including K-means, fuzzy c-mean and Expectation Maximization. One of the advantages regarding these methods is that they consume less time, because do not use training data. However, these algorithms are sensitive to noise and intensity inhomogeneities since do not take into consideration the spatial information.

K-means cluster algorithm is the process of classifying a given data set through a certain number of K-clusters. The process is based on the iteractively computation of a

mean intensity for each class and the segmentation is done by classifying each pixel in the class with the closest mean. The mean of each class is iteratively updated as new pixels are added.

The fuzzy c-means algorithm generalizes the K-means algorithm, which tries to minimize the intra-cluster variation through iterations. Instead of classifying a pixel into a fixed cluster, each pixel can potentially belong to multiple clusters, based on the probability of belonging to each cluster. This algorithm provides a softer segmentation than the K-means.

The Expectation Maximization method iteractively calculates the maximum-likelihood estimation of the means and covariance. This algorithm is composed by two steps: calculation of the expectation of the likelihood and then its maximum. Although these methods are classified as unsupervised, they do need some initial parameters. The EM algorithm has demonstrated greater sensitivity to initialization, when compared to K--means or fuzzy c-means algorithms [9], [48].

2.2.5 Deformable models

Based on the prior knowledge of geometry and physics, the deformable models are able to solve segmentation problems in a set of images over time and from different individuals. Level sets and active contours are examples of deformable models, which will be discussed in this section.

The level set is a numerical method used to account for the evolution of contours and surfaces. A curve is represented as the level set of a 2D scalar function, defined on the same domain as the image. The set of points that have the same function value defines the level set. The level set method evolves a curve by updating the level set function at fixed coordinates through time. Thus, in a typical approach, the contour is initialized by a user and then it is evolved until it fits the form of the anatomical structure intended [49].

Active contours, also called snakes, are based on the minimization of energy in splines. By defining an initial contour or a seed inside the object of interest, the active contour algorithm attempts to minimize the contour energy by moving the contour inside the target object.

2.2.6 Atlas-based

AB segmentation algorithms use pattern recognition techniques to assist segmentation. By assigning a relationship between segmentation labels and image intensities in a particular atlas, the atlas-guided algorithm promotes an association between the pixels or voxels of an unlabeled image and the segmentation labels [9], [50]. The segmented labels in the atlas can be done manually or by one of the approaches discussed previously.

Simple AB segmentation is based on a probabilistic atlas, where all the images available are summarized. First, all images available are co-registered in a single atlas coordinate frame and statistics about the labels are pre-computed in the atlas space. The co-registered images are averaged to originate only a template image. To segment a new image, it is registered with the template and then it is segmented based on the segmentation of the template.

Multi-atlas segmentation is an alternative strategy, in which different atlases can be used for the segmentation of the novel image, depending on the criteria. This method can be divided in four major steps: generation of atlases, registration, label propagation and label fusion (figure 2.3).

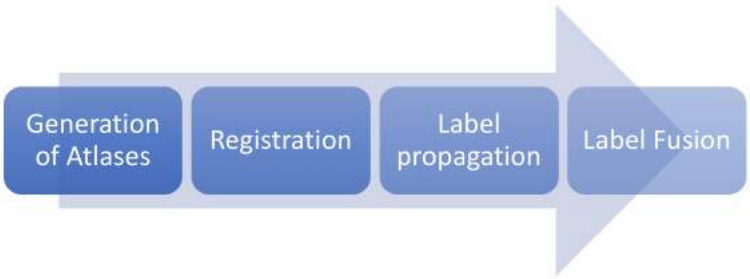


Figure 2.3: Schematic representation of Multi-Atlas Segmentation steps.

The first step is the generation of atlases, i.e. the choice of which labeled training images yield the maximum performance when new data are segmented. This preselection can be done manually, by visual inspection or automatically. Accuracy of segmentation is affected by the proper choice of the training data, so low-quality images should be discarded. However, although it reduces the computational time, reducing the number of atlases can also affect segmentation.

After choosing the atlases, a match must be stablished between each atlas image with the target image - registration. Typically, it is computed between each atlas and the novel image, one independent intensity-based registration. Nevertheless, the choice of the registration approach depends on the data to be analyzed and can be constituted by the different modalities discussed in 2.1.

Once the significant atlases are selected and spatial correspondence is stablished between each atlas and the novel image, the next step is to propagate the atlas labels to the novel image coordinates. This can be achieved by nearest neighbor interpolation, where a single label is transferred from the atlas to each image pixel/voxel, linear interpolation, signed distance maps or learning algorithms.

Finally, the atlas labels propagated to the target image are fused to obtain the final segmentation result. The simplest fusion methods used are best atlas selection and majority voting. In best atlas selection, a single atlas is chosen based on similarity measurements between the registered atlas and novel image intensities. The disadvantage is that ignores the information in all other atlases. Majority voting uses the information of labels from

all atlases, to choose for the most frequent label at each location. However, it has the drawback of discarding the image intensity information. Alternatively, in weighted voting, each atlas is associated with a weight. The higher the similarity between the target image with the atlas, the higher the correspondent weight.

Additional information about multi-atlas segmentation can be found in the survey of J. E. Iglesias & M. R. Sabuncu [50]. Several examples of different approaches for multi-atlas segmentation of the heart, mitral valve, rat muscles and human abdominal multi-organ are available in [51, 52, 53, 54].

Chapter 3

Magnetic Resonance Imaging and Diffusion

3.1 MRI Principles

This chapter was mostly based on the literature published by McRobbie & Donald W. *et al* [55].

MRI is a non-invasive diagnostic technique based on the atomic nuclei magnetic properties and the interaction of a nuclear spin with an external magnetic field, B_0. MRI provides access to the anatomy and physiologic processes of the human body, with a high spatial resolution and excellent soft tissue contrast. Hydrogen is the most commonly used molecule due to high sensibility and abundance in human tissues. Hydrogen's proton have an intrinsic magnetic moment, μ, and when subjected to a magnetic field B_0 it rotates at a certain frequency, proportional to the field strength, which results in an angular momentum, the spin.

In the absence of an external magnetic field, the rotation of protons is random, and so the net value of the magnetization is null. On the other hand, when applying an external magnetic field, protons will align in the direction parallel or antiparallel to that of the field. Besides having a rotation movement around the magnetic field vector, the nuclear spin also rotates around that direction. The existence of two energy levels that protons can occupy inside the magnetic field is responsible for the two possible directions of alignment, due to the Zeeman effect. There is a division of the degenerated energy level, into a state of high energy (antiparallel) and low energy (parallel). Based on the Boltzmann distribution, the configuration of low energy is the preferred one. The difference of the protons' distribution between the levels is the true contribution for the MRI signal, called longitudinal magnetization M_0.

The frequency at which a proton rotates in the presence of an external magnetic field

with intensity B_0 is given by

$$w_0 = \gamma B_0 \tag{3.1}$$

where w_0 is the Larmor frequency, γ is the gyromagnetic constant ration between the magnetic and the angular moment and B_0 is the magnitude of the magnetic field.

In order to acquire signal, there must be transitions between the higher and lower states of energy (antiparallel- and parallel), which must be induced by an external energy source, the Radiofrequency (RF) pulse. The RF pulse must be applied perpendicular to B_0 and at the Larmor frequency of the element in study so to induce resonance, which correspondes to the gap between the two levels of energy [56].

At rest, there is no transversal magnetization and the net magnetization vector only has the z component. When applying the RF pulse, the z component is reduced from its equilibrium value, M_0, and the transversal xy component becomes non-zero. When this is done, the spins become synchronized and rotate at a given angle, which depends on the duration and intensity of the pulse. When the RF pulse is turned off, only the external magnetic field is on, so the spins relax into it again.

Considering the application of a 90° RF pulse, the longitudinal component $M_z(0)$ equals to zero and the transversal component $M_{xy}(0)$ has its magnetization vector arbitrary. The magnetization at later times is given by the Bloch equations,

$$M_z(t) = M_0(1 - e^{-t/T_1}) \tag{3.2}$$

$$M_{xy}(t) = M_{xy}(0)(e^{-t/T_2}) \tag{3.3}$$

where T1 is the longitudinal relaxation time and T2 the transversal relaxation time. T1 is the relaxation along the B_0 direction, also called spin-lattice relaxation since it corresponds to the re-establishment of the thermal equilibrium in the local environment. It is also defined by the time taken for 63% of M_0 to recover after a 90° RF pulse. T2 is the relaxation along the plane perpendicular to B_0, or spin-spin relaxation, the time that transversal magnetization M_{xy} takes to fall 37% of its original value, determined based on the RF pulse duration and intensity.

By plotting both curves on the same graph (figure 3.1) for a tissue with T1=5xT2 the differences of time scales are well distinguishable. Intrinsic magnetic design and differences in magnetic susceptibilities between different tissues cause spatial variations in the strength of the magnetic field, which influence the transversal relaxation. These interactions between spins and field inhomogeneities also contribute to T2, becoming a $T2^*$ rate instead, shorter than T2.

The Magnetic Resonance (MR) signal is obtained based on the Faraday's Law of Induction, wherein a changing magnetic field induces a voltage in a nearby conductor. In this case, the variation of M_{xy} is detected by a receiving coil, which induces the generation of electric current. The signal obtained by the coil is called Free Induction Decay (FID).

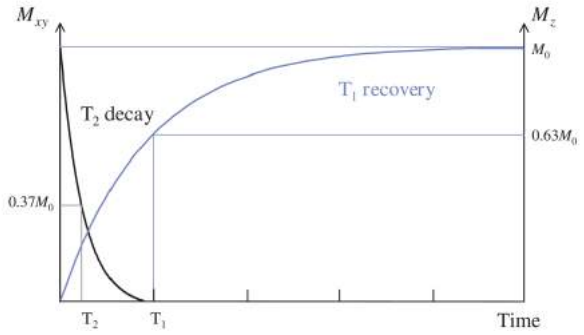


Figure 3.1: T1 and T2 relaxation times. Although occurring at the same time, T2 is faster than T1 [55].

3.1.1 Image Contrast

In order to emphasize certain tissues, MRI images can be weighted in three different parameters: T1, T2 and Proton Density (PD). It is important to realize that these three parameters are properties of a given tissue and that an image can be obtained based on that certain property.

T1 weighted (T1w) images are obtained by setting a short time between two excitation RF pulses, the Repetition Time (TR) (figure 3.2-1A). This allows less time for the net magnetization vector to recover, which means that long T1's tissues do not have time to relax completely, weakening the signal. On the other hand, tissues with a short T1 (shorter that TR) have time to relax completely, recovering their longitudinal magnetization prior to being flipped by the second 90° RF pulse. This results in a strong signal. On the contrary, selecting a long TR value reduces the T1 contrast between tissues, since they have time to recover their magnetization (figure 3.2-1B).

To measure the signal, it is necessary to apply a 180° RF pulse after the 90° one, to realign the spins. After the first 90° RF pulse, FID occurs. Then, following the 180° RF pulse surges a spin echo (figure 3.3). The time between the 90° pulse and the echo is called echo time (Echo Time (TE)). If a series of 180° RF pulses is applied after the 90° RF pulse, T2 decay can be measured by the curve that passes by the maximum of the FID and following echos.

T2 weighted (T2w) images are obtained by controlling the TE. If a short TE is used, the transversal magnetization of the tissues does not have time to relax completely, thus resulting in a poor contrast (figure 3.2-2A). However, by setting a long TE value, relaxation has time to occur for both long and short T2 tissues (figure 3.2-2B) [57].

Lastly, PD weighted (PDw) images quantify the protons that exist in a certain tissue. PDw images are obtained by controlling TR and TE to reduce T1 and T2 effects. On one hand, TR needs to be long enough to allow a complete longitudinal relaxation of the spins, diminishing the T1 differences between tissues. On the other hand, TE should be short enough not to allow T2 contrast to develop, reducing the effects of the transversal relaxation. Therefore, signal contrast is obtained by the relative PD between tissues.

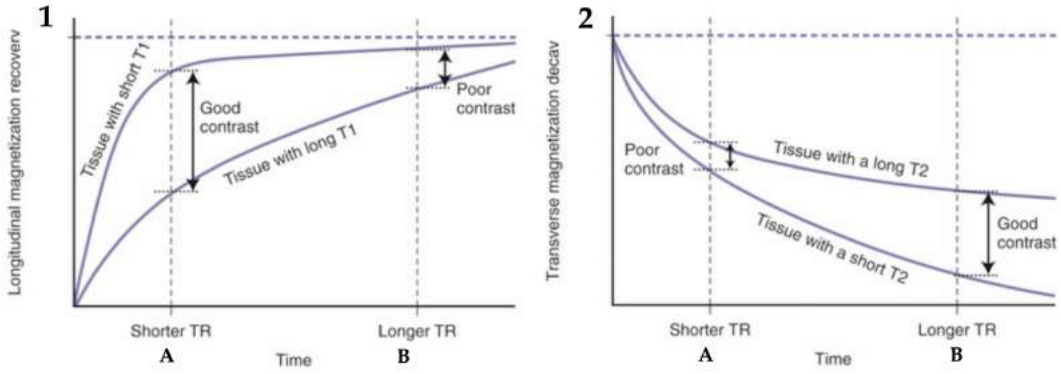


Figure 3.2: T1-weighted tissues with different T1 relaxation times (1) and T2-weighted tissues with different T2 relaxation times (2). A good contrast between T1w tissues can be obtained by setting a short TR, since the magnitudes of their longitudinal magnetization recovery will be different. A smaller difference between their recovered magnetization vectors is found when TR is long, since they have time to recover their longitudinal magnetization, resulting in poor contrast between them. A good contrast between T2w tissues can be obtained by setting a long TE, allowing almost complete transverse magnetization recovery. By setting a short TE, there is almost no difference between the loss of transverse magnetization of the two different tissues. Adapted from [57].

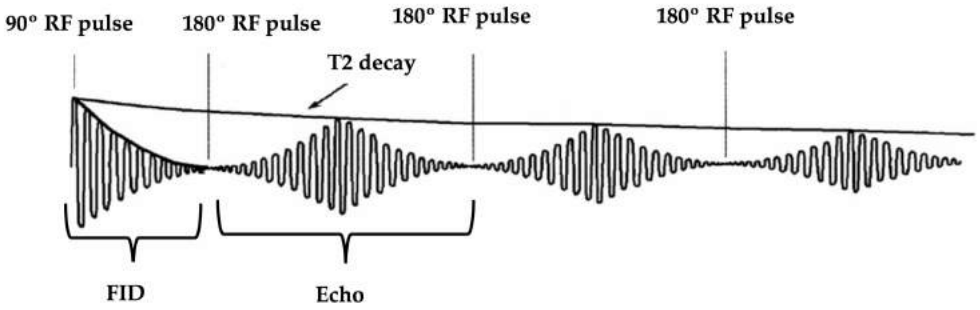


Figure 3.3: T2 decay curve. After the 90° RF pulse, FID occurs. T2 decay is defined by the curve that passes the maximum of the FID and following echos, result of setting the 180° RF pulses [58].

Contrast between tissues or even signal intensity of the same tissue can have different graylevels distribution across the same image due to tissue heterogeneity or due to field inhomogeneities [59].The inhomogeneities present in the magnetic field of the MRI machine are responsible for creating an undesirable low frequency smooth signal - bias field - that contaminates MRI images (T1w, T2w, PDw, DWI ...). This bias field reduces the high frequency content of an image, such as edges and contours, changing the intensity value of pixels across it. Although it may not have a direct clinical diagnosis impact, it has a tremendous influence on imaging processing. Registration and segmentation based on intensity (e.g. edge-based segmentation) are highly affected by the bias field, since can wrongly classify a region. Thus, MRI images require a pre-processing step before extracting meaningful information, to correct for the effect of the bias field.

3.2 Diffusion Weighted Imaging

Molecular diffusion, or Brownian motion, is the random movement of molecules in fluid (e.g. water) drove by thermal energy. Thus, in the three-dimensional space, the water molecules' trajectory is not predictable. When restricted to a close space, the molecules' trajectory is no longer random since the physical barriers restrain the natural process of diffusion. The displacement can be characterized by the diffusion constant, D, given by the Einstein's equation,

$$D = \frac{\overline{R^2}}{6t} \tag{3.4}$$

where $\overline{R^2}$ is the mean square displacement and t the time of observation, at a constant temperature [60]. The idea behind Einstein's formalism is based on the experience of measuring the individual displacement of a given number N of labelled water molecules in water, after a given time interval Δ. For each displacement distance r, the number n of water molecules that reached that distance are counted. Then, a histogram of the relative number of labelled molecules (n/N) versus displacement distance (r) is plotted. Figure 3.4 shows a typical displacement distribution of diffusion in a homogenous medium, described as having a Gaussian distribution.

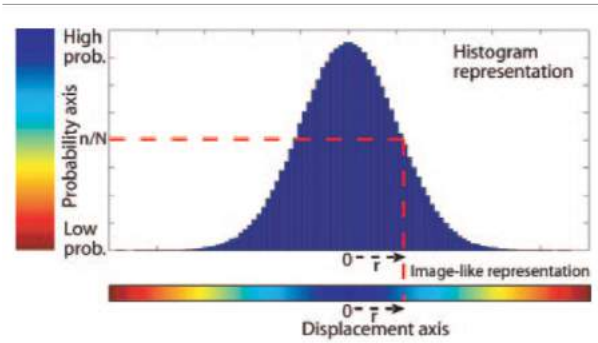


Figure 3.4: Typical displacement distribution due to diffusion in a one-dimensional model. For each displacement distance r, there is a corresponding proportion of molecules (n/N) within a voxel that were displaced that distance at a given interval Δ (the duration of diffusion experiment). For example, the red line indicates that at a given distance r, a certain proportion of molecules traveled that given distance. The horizontal color bar also shows the same Gaussian distribution and is indicative of high (blue) and low (red) probability of displacement. Adapted from [61].

In an unrestrained environment, there is no particular directionality in the water movement in a given amount of time, so it is considered as isotropic diffusion. However, in contrast to that environment, biological tissues are highly heterogeneous. They are composed by various compartments and barriers of different diffusivities, so the movement of water molecules is restrained by compartmental boundaries and other molecular obstacles, such as cell membranes and walls, fibers, axonal myelin sheaths and macromolecules. In this case, the diffusion distance is reduced when compared to unrestricted

diffusion. This phenomenon is called anisotropic diffusion, since it is time and direction dependent [61].

Diffusion is measured by the ADC, which depends on the gradients and time of application. The term *apparent* is due to the impossibility of differentiating diffusion from other sources of water mobility in in vivo acquisitions. ADC can take values from 0 to D, varying from absence of diffusion to be the only water motion phenomenon present, respectively.

DWI allows mapping of the diffusion process of molecules in tissues, non-invasively, generating images with high contrast between tissues and with a micro structural resolution. Contrast between tissues in T1 and T2 weighted is given by changes in the relaxation time, while functional MRI lies on blood oxygen level dependency. On the contrary, contrast between tissues in DWI is given by the changes of the water diffusion, dependent on the temperature, in biological tissues, namely inter, intra and extracellular. Consequently, these functional changes are visible before identified alterations on the morphologic routine sequences.

DWI signal, S_i, can be described by equation 3.5,

$$S_i = S_0 e^{-bADC_i} \tag{3.5}$$

where i is the direction to which gradients were applied, S_0 is the signal intensity with no diffusion weighting, ADC_i is the apparent diffusion coefficient measured in i direction and b-value (s/mm^2) defines the sensitivity degree to diffusion phenomenon and determines the strength and duration of the diffusion gradients. It can be defined by equation 3.6.

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) \tag{3.6}$$

Equation 3.6 shows the dependence of the gyromagnetic constant γ, amplitude of the diffusion gradient G (mT/m), the duration of each gradient δ (ms) and the time interval between gradient pairs Δ (ms). Manipulating these parameters allows different weightings of diffusion.

In order to fit the exponential function to calculate the ADC map, one must measure at least two b-values. Multiple b-values can be used to calculate the ADC map, improving its accuracy. However, it increases the scanning time [62]. Usually, are chosen a b-value of 0 s/mm^2 and one of 1000 s/mm^2 to supress normal background, depending on the organs studied. A b-value of zero results in a T2-weighted image, as an anatomical reference, where healthy tissues are more attenuated than lesions. The higher the b-value, the stronger the diffusion weighting and so the higher the contrast in pathogenic regions. ADC can so be calculated based on equation 3.7, where b_0 is the lower value and b_1 the higher value of b. The more random the movements of molecules, the more dispersed are the spin's phases and so the signal intensity will be lower. In this case, there are low restrictions to diffusion and the region will appear hypointense. Therefore, a high

value of ADC will correspond to a low signal intensity in DWI images. Thus, ADC is a marker of the tissue density and cellularity. High cellularity tissues or swelling result in a contraction of the extracellular space, causing restricting diffusion, as indicated by a low ADC. On the other hand, low density tissues exhibit an increase ADC. ADC together with signal intensity in high b-value images are good indicators of the biophysical properties of skeletal metastases.

$$ADC_i = -\frac{\ln \frac{S_i}{S_0}}{b_1 - b_0} \tag{3.7}$$

The signal intensity of DWI images depends on the TE, diffusion coefficient, b-value chosen and proton density. The TR normally used is long (5000-15000 ms), so there are no contributions from T1. However, the TE used is short (60-100 ms), thus T2 contamination may be observed. This can result in misinterpretations on DWI, due to the contribution of water proton diffusitivy and intrinsic tissue T2 relaxation time, since cellular disease and tissues with long T2-relaxation times (e.g. cystic areas, necrosis, fluid, pleural effusion) can appear with a similar intensity on DWI [63]. Thus, high signal intensities in DWI do not always relate with diffusion-restricted areas. This phenomenon is known as the T2 shine-through effect. The same happens with tissues with short T2 relaxation times, whose intensity in DWI might be erroneous low (T2 dark-through effect) [64, 65].

T2 related effects in DWI can be eliminated by acquiring and relating with ADC maps. However, these maps have low tissue contrast when compared to normal DWI, hindering lesion detection. Instead, one can acquire images with high b-values, while preserving tissue contrast. Images acquired with high b-values show a lower signal in unrestricted tissues (low ADC) than in diffusion-restricted tissues (high ADC), although signal intensities decrease in all tissues. Thus, it increases contrast between diffusion-restricted and unrestricted tissues, while decreasing the relative contribution of T2-relaxation time [66]. Nevertheless, acquiring images with high b-values is a challenging task due to their intrinsically low Signal to Noise (SNR) ratio and proneness to severe eddy current distortions from the gradients used. computed Diffusion Weighted Imaging (cDWI) has been proposed by Blackledge *et al* [67] as a mathematical computation technique, based on equation 3.7, that extrapolates images with high b-values using at least 2 lower b-values (0-1000 s/mm^2). Typical b-values of 0 and 1000 s/mm^2 are used to calculate tissue ADC maps, as was discussed previously, allowing to compute higher b-value images. These computed images are less prone to distortion and poor SNR ratio , which can improve diagnosis sensitivity and specificity.

The typical imaging sequences used to acquire a DWI signal are Pulsed Gradient Spin Echo (PGSE) and Single Shot Echo Planar Imaging (SS-EPI).

3.2.1 Pulsed Gradient Spin Echo

This sequence was created by Stejskal and Tanner [68], being composed by two RF pulses of 90° and 180°, and two symmetric magnetic gradients with intensity G, before and after the 180° RF pulse, as seen in figure 3.5.

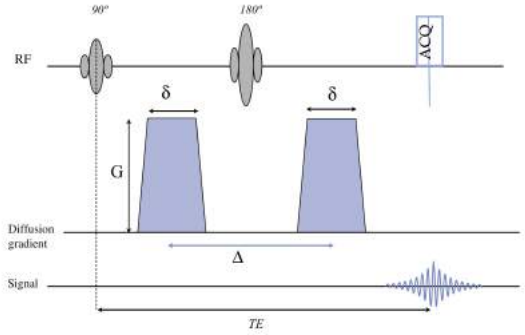


Figure 3.5: PGSE sequence for DWI, always preceded by a spin-echo EPI acquisition. δ symbolizes the pulse width, Δ the center-to-center spacing, G the magnitude of the gradient and ACQ is short from acquisition. Adapted from [55].

The 180° RF pulse inverts the phase wrap imposed by the first gradient, while the second gradient resets the phase by the same angle initially imposed by the first gradient. Spins which are stationary will not be affected by the diffusion gradient pair since any phase accumulation due to the first gradient will be reverse by the second. However, applying the two gradients to diffusing spins will result in a loss of coherence, since they move into different locations between the two gradients (figure 3.6). Thus, these spins will fall out of phase, resulting in an attenuation of the signal.

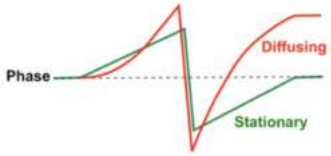


Figure 3.6: Loss of phase coherence of an individual diffusion spin. In green, a stationary spin is not affected by the paired gradients. In red, a diffusing spin is affected by the gradients and so is dephased [69].

3.2.2 Single Shot Echo Planar Imaging

SS-EPI is composed by one RF pulse of 180°, between two opposing polarity diffusion gradients. After the 180° RF pulse, a series of interchangeable polarity gradients are applied, forming a train of gradient echoes. Each echo is phased encoded differently, in order to fill the k-space. Unlike PGSE, where only one line of imaging data is collected within each TR, in SS-EPI, multiple lines can be acquired after a single RF pulse. By manipulating the intensity of the series of gradients applied, TE can be reduced and diffusion measurements can be obtained [55, 70].

SS-EPI is the most used sequence to acquire diffusion weighted signal, due to the shortening of MRI times. This is a great advantage, when compared to other sequences, since artefacts due to patients' movements during the acquisition time are less probable, as the acquisition time is short.

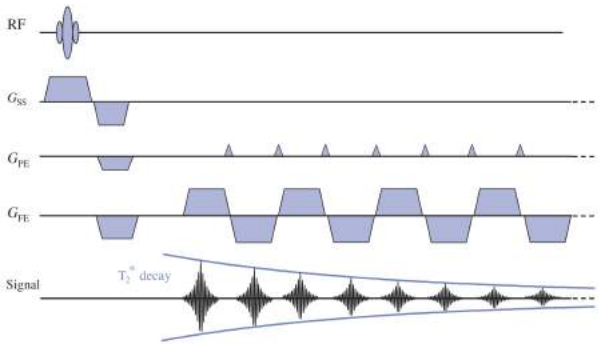


Figure 3.7: SS-EPI sequence. First, a 180° RF pulse is applied between two diffusion gradients. Phase coding is given by rapid inversions of small phased-gradients (G_{PE}), placed between two inverted frequency encoding gradients (G_{FE}). This generates a train of echoes (Signal), that fills the k-space rapidly and easily. Adapted from [55].

Chapter 4

Oncologic Applications to study Multiple Myeloma

There is no clear agreement on the accepted methods for assessing tumor response in skeletal sites with metastatic disease or local destruction. Treatment response can be assessed by a combination of imaging techniques, serum and urine biochemical biomarkers and clinical evaluation [71].

DWI is used to distinguish and describe benign and malignant lesions, enabling tumor staging, foreseeing treatment outcomes, evaluate treatment responses and relapses. The clinical evaluation can be performed by selecting regions of interest placed on the tumor or parts of it and measurements of mean ADC values, or voxel-by-voxel comparison of changes in ADC, when assessing treatment response. DWI has high value for cancer staging and an increasing value for detecting metastatic disease in cases of breast, prostate, female pelvis, bladder and kidney tumors, and also in cases of hematologic cancers that involve bone infiltration [64].

4.1 Multiple Myeloma

MM is the second most common haematological neoplastic disease, accounting for 10% of the haematological malignancies and 0.8% of all new cancer cases. MM corresponds to an advance stage condition in the group of monoclonal gammopaties and is characterized by an abnormal proliferation of malignant plasma cells throughout the bone marrow, resulting in an increased concentration of monoclonal paraprotein (M protein). It is often associated with Hiper**C**alcemia, **R**enal insufficiency, **A**nemia and lytic **B**one lesions (CRAB criteria for MM) and immunodefiency [72, 73].

In early biological stages this disease is asymptomatic, being classified as Monoclonal Gammopathy of Unknown Significance (MGUS), with a prone progression to Smoldering

Multiple Myeloma (SMM) and finally to symptomatic MM. MGUS is present in 3-4% of the population over the age of 50 years [74], and its rate of progression to MM is 0.5-1% per year. The rate of progression depends on the concentration and type of M protein, serum free light chain ratio, bone marrow plasmocytosis[1], proportion of phenotypically clonal plasma cells, cytogenetics alterations and presence of immunoparesis[2] (figure 4.1) [75].

MM pathogenic mechanisms have been recently more clarified and include several molecular and genetic transformations occurring over the years. Once the malignant plasma cells are formed, they interact with hematopoietic and stromal cells in the bone marrow, resulting in a disturbance of the homeostasis between cells and within the extracellular matrix. Subsequent to these interactions, there is an abnormal increase in the number of these clonal plasma cells and activation of several signal cascades that support its proliferation, survival, migration, drug resistance and tumor angiogenesis. In addition, there is activation of osteoclasts that causes osteolysis, damaging the bones. The bones that are most affected are localized in the skullcap, spine, pelvis, costal grid and proximal long bones, as the femur.

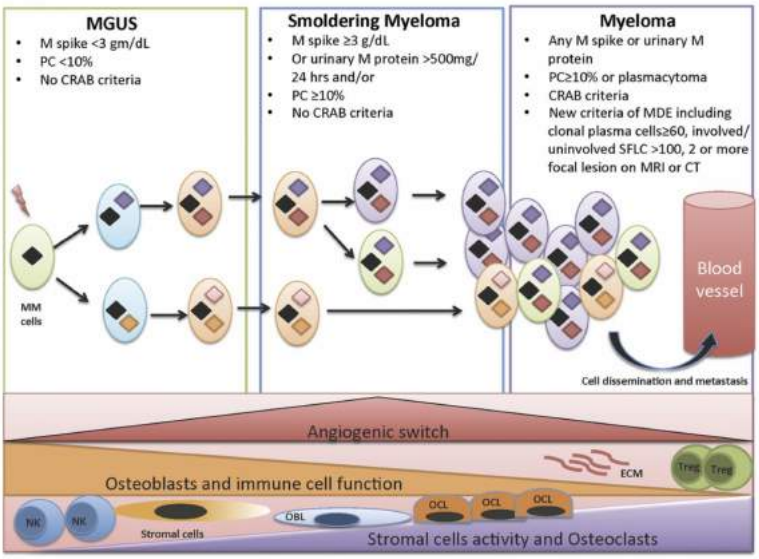


Figure 4.1: Biological evolution from MGUS to SMM to symptomatic MM and clinical criteria summary. This requires clonal evolution and heterogeneity, which is not only cell autonomous but also dependent on the interactions of the tumor cells with the bone marrow microenvironment. This includes immune cells such as T-regulatory cells (Tregs), myeloid derived suppressor cells (MDSCs), natural killer (NK) cells, osteoclasts (OCL), osteoblasts (OBL), angiogenesis, and marrow stromal cells (MSCs) [76].

MM diagnosis includes blood and urine tests to assess and quantify the monoclonal protein M and to assess organ involvement, biochemical and histopathology tests as bone

[1] Increased number of plasma cells.
[2] Reduction in the levels of polyclonal or uninvolved immunoglobulins.

marrow aspirate assessment and biopsy to calculate the percentage of clonal plasma cells and medical imaging [77].

Until recently, the IMWG stated that Whole Body X-ray (WBXR) was the gold standard for the evaluation of MM bone disease. However, it has been shown that its detection limit is low. For instance, to detect an osteolytic lesion, 30% to 50% of the trabecular bone has to be reabsorbed. Furthermore, it fails at depicting the cause of painful lesions and osteoporosis in MM, has low visualization of the spine and pelvis and lacks at detecting small lesions [75]. Imaging technology has evolved in the last decade, wide spreading the use of MRI and its functional possibilities, Whole Body Low Dose Computed Tomography (WBLD-CT) and F18-Fluorodeoxyglucose Positron Emission Tomography (F18-FDG PET) to assess bone disease and bone marrow infiltration in MM.

IMWG sustains that MRI is the gold standard for detection of bone marrow involvement in MM, since it has shown higher sensitivity than WBXR and detects bone marrow focal lesions long before the development of osteolytic lesions, seen on WBXR. WB-MRI is preferred since MRI of the spine or pelvis can ignore some lesions. Conventional MRI protocols for MM include T1-weighted, T2-weighted with fat suppression, in- and opposed-phase imaging and contrast-enhanced T1-weighted sequences.

The typical MRI patterns that represent bone marrow involvement in MM are normal bone marrow pattern, focal, diffuse, combination of focal and diffuse pattern and a variegated or "salt and pepper"pattern. Focal patterns, present in 30%-50% of MM patients, are localized zones of myeloma cell infiltration with a diameter 5mm or greater. Diffuse patterns occur when there is almost complete bone marrow replacement by the myeloma cells, which happens in 25%-40% of MM patients. The combined form of focal and diffuse pattern has an incidence of 10% in this population, while the normal bone marrow pattern has 15%-25%. Although less common (1%-5%), the variegated pattern develops multiple small bone marrow focal lesions [78]. Figure 4.2 shows examples of appearances of focal, diffuse and variegated pattern on T1-weighted images.

WBLD-CT protocol is the preferred recommended method for detection of lytic lesions in MM, since it enables small osteolytic lesions detection with less radiation exposure [79] and it is a very easy and fast technique. However, F18-FDG PET-CT also exhibits high value for the identification of lytic lesions and extramedullary masses, as well as distinguish active from inactive metabolic lesions due to the combination of functional and morphological information. For this reason, recent consensus recommends the use of F18-FDG PET-CT to distinguish between SMM and MM if WBXR is negative and WBCD-CT and WB-MRI is not available [80].

Figure 4.3 depicts the European Myeloma Network flow chart for the recommended imaging technique, based on, e.g. the presence of lytic lesions, number of focal lesions, risk of fracture.

Although WBXR was widely used as the preferred imaging technique to study MM, it was substituted by WBLD-CT to study lytic bone lesions. However, both WBXR and

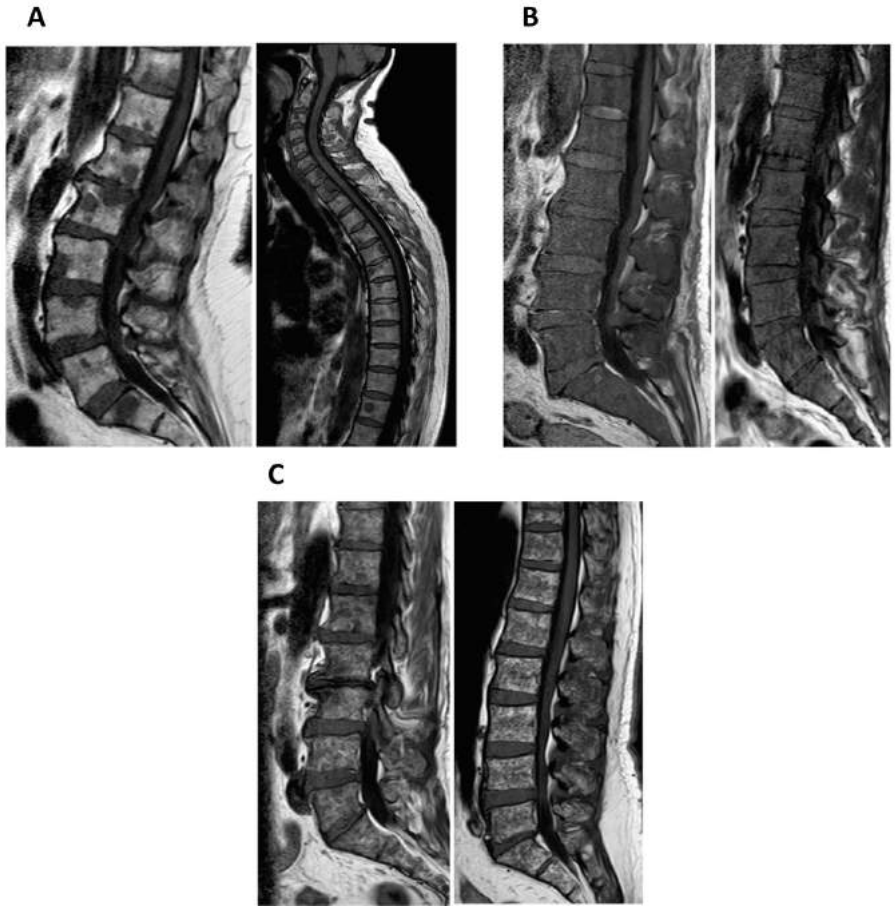


Figure 4.2: Appearances of focal, diffuse and variegated pattern on T1-weighted images. (A) Focal pattern: On T1-weighted images, focal lesions are darker than yellow marrow and slightly hypointense or isointense to intervertebral disc and muscle. (B) Diffuse pattern: On T1-weighted images, the abnormal process replaces the normal bone marrow signal, and the diseased marrow appears darker or isointense to the intervertebral discs. (C) Variegated pattern: On T1-weighted images, the bone marrow is very inhomogeneous with innumerable small lesions [26].

WBLD-CT cannot assess bone involvement by the myeloma cells and response to treatment. Bone metabolism is slow and lytic lesions, when present, do not seem to disappear over time, even in the cases where there is a response to therapy.

MRI has been commonly used to perform structural studies in MM patients. However, these structural studies can only infer about tumor size measurements, having limitations for assessing therapy response. More recently, functional techniques included in MRI, as Dynamic Contrast Enhanced (DCE)-MRI and DWI-MRI , start to be explored in these population of patients giving information on quantification of bone response [82]. These functional techniques complement the structural studies by adding information about the bone marrow cellularity and vascularization, improving the overall performance of MRI .

F18-FDG PET-CT uses a quantitative method to measure tumor load that is the uptake

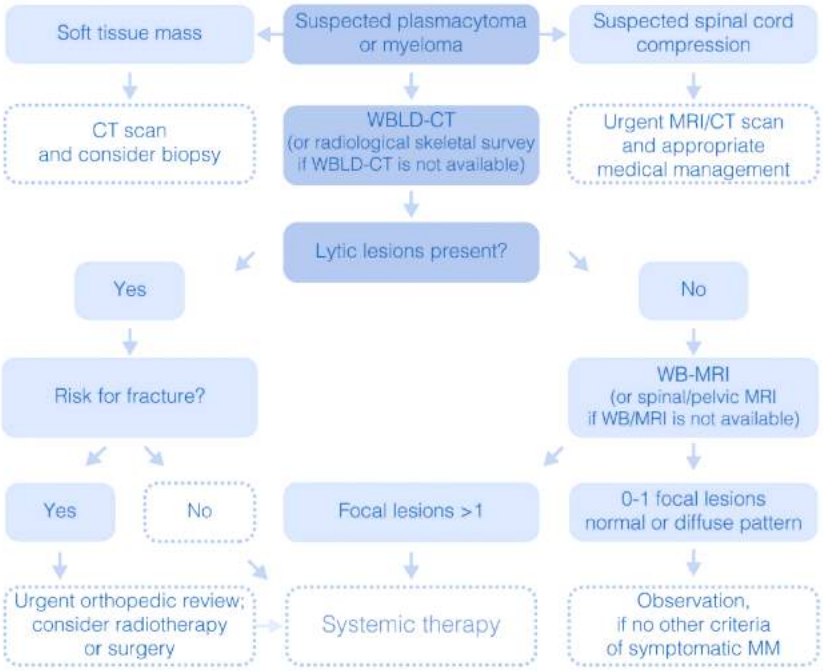


Figure 4.3: Algorithm for imaging in MM. In the case of a plasmocytoma (soft tissue mass due to plasma cells growing), a CT of the area and a needle biopsy are needed. If spinal cord compression is verified, an urgent MRI or CT is obligatory to assess the best care (radiotherapy or surgery if there are bone fragments in the spinal canal). In the case of myeloma, WBLD-CT is recommended for the evaluation of the skeleton. In this case, if lytic lesions are present, the patient is classified as symptomatic and needs systematic therapy. If not, a WB-MRI is performed and if more than one focal lesion is found, the patient is classified as symptomatic. Adapted from [81].

values of F18 Fluorodeoxyglicose (FDG) that is the Standard Uptake Value (SUV) scale. Though, it is highly influenced by the serum glucose level, as well as uptake time, scanner calibration and reconstructive techniques.

With the incorporation of DWI to the MRI protocol, WB-DWI offers ADC measurements to quantify disease burden, with lack of contrast or ionizing radiation, assessing skeletal complications and differentiating between normal and pathological marrow. Although being also influenced by scanner manufacturer and field strength, ADC is considered to be less prone to equipment and physiological effects [83]. Furthermore, preliminary work has demonstrated high repeatability in measurements of ADC of MM patients who underwent WB-DWI twice (2,8% of mean coefficient variation), confirming the reliability for quantifying treatment response in myeloma [84]. DWI can be used to assess bone marrow infiltration for diagnosis and monitoring of treatment response. The use of DWI protocols is very recent and its usefulness in MM is still under investigation, although active research is being conducted in this area [85, 86, 87].

4.2 Magnetic Resonance Imaging of Multiple Myeloma

4.2.1 Bone marrow reconversion imaging

At birth, the skeleton of a normal human being is filled with red bone marrow, which is composed of hematopoietic cells (water 40% and protein 20%) and fat cells (40%). Red marrow is converted to yellow bone marrow throughout an individual's life, becoming more prevalent with age. The yellow marrow is composed by 80% fat, 15% water and 5% protein. The conversion starts from the epiphysis limbs and disperses towards the articular ends, progressing to the axial skeleton over time [88].

The differences between water and fat content in yellow and red marrow produce different signal intensities on T1w spin echo MRI: yellow marrow has an hyperintense signal, contrasting with the red bone marrow, which shows an hypointense signal.

Yellow bone marrow has the ability of reconverting to red bone marrow when there is an increased need of hematopoietic cells. Among non-medical conditions, for instance heavy smoking, doing sports with a large oxygen debt (e.g. long-distance running, free diving) are considered to increase the need of hematopoietic cells. Obesity and related respiratory disorders, diabetes, recovering anaemia, inflammatory diseases and some kinds of treatment (e.g. post-chemotherapy, administration of hematopoietic growth factors) are considered some of the medical conditions that may lead to a growing of hematopoietic cells [89].

As was already stated, MM is characterized by an abnormal proliferation of malignant plasma cells throughout the bone marrow. This infiltrative bone marrow pathology shows high cellularity patterns and water content, while decreasing the amount of fat. As a result, show hypointense signal on T1w spin echo MRI, being easily distinguishable from yellow marrow from surrounding tissues.

In the first days after treatment with chemotherapy or radiotherapy, the bone marrow appears hypointense on T1w images, hyperintense on high b-value DWI with an increased ADC, since undergoes cellular death. On the following days, there is yellow marrow conversion, reducing the overall ADC and signal intensity on high b-value DWI. Several weeks into treatment, hematopoietic recovery occurs, decreasing the signal intensity on T1w of the bone marrow. This may be a major problem when analyzing T1w MM images, since it may be very difficult to distinguish between a MM pattern from hematopoietic marrow due to reconversion [63].

Although there is no clear agreement on distinguish MM lesions from hematopoietic marrow due to reconversion in MRI, - it is not feasible to perform biopsies in all lesions - MM lesions can be defined as having [29, 88, 90, 91]:

- T1w signal intensity equal to or lower than that of muscle or nondegenerated intervertebral discs;
- Hyperintense signal on short TI inversion-recovery images;

- Less than 20% signal intensity decrease on opposed-phase images;
- More than 40% peak enhancement on post-contrast T1w, when compared with that at baseline;
- Hyperintense signal on DWI images (introduced very recently, not included in the IMWG 2014 guidelines [75]).

In asymptomatic MM patients, new IMWG 2014 guidelines [75] regarding imaging assessment stipulate more than one focal lesions (> 5 mm) as an indication to be considered to start treatment. Diffuse infiltration is often associated with a worse prognosis and advanced disease. This is perhaps due to the absence of an agreement on classifying a diffuse pattern on conventional MRI, which is often a subjective diagnosis. The recent introduction of DWI protocols and integration of ADC measurements has potential to reduce this subjectivity.

Figure 4.4 shows the differences in signal intensity on high b-value images and ADC during the disease course from MGUS, over SMM to MM. The early and later changes in response to treatment are also displayed. Yellow marrow is mainly composed by fat, showing restricted diffusion patterns. MGUS signal is very similar to that of yellow marrow, since has low concentration of MM cells. With the evolution of the disease, there is an increase of the MM cells. The bone marrow infiltration has to be high enough to result in a decrease in fat cells and so be noticeable both on conventional MRI and DWI. The figure also indicates that there is a signal difference between hematopoietic marrow and the several forms of MM, suggesting that there is a threshold that could distinguish red marrow from myeloma infiltration patterns.

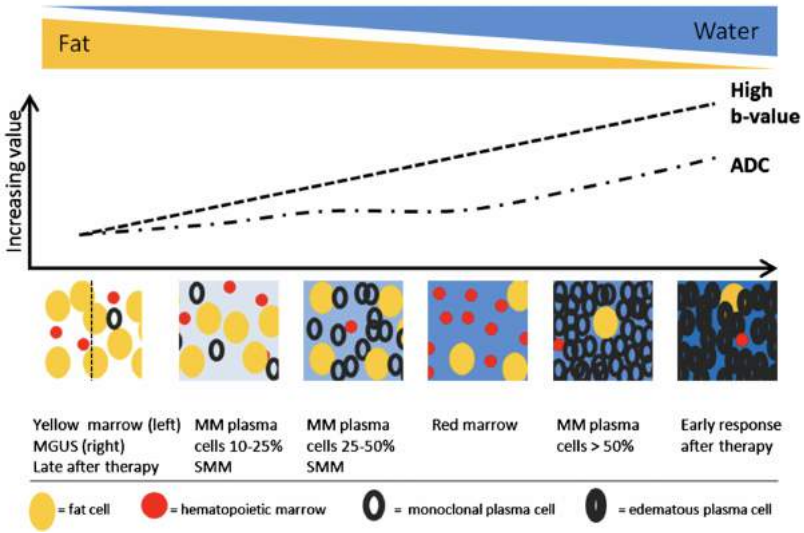


Figure 4.4: Signal intensity change on high b-value and ADC images during the MM disease course: from MGUS, over SMM to MM. The percentages of M cells are also indicated. Changes early and late after therapy, normal yellow and red marrow can be compared with the MM stages of disease [63].

Chapter 5

Materials and Methods

5.1 Study Design and Population

As part of the common clinical practice for the study of neoplastic diseases in the Champalimaud Clinical Centre (CCC), WB-MRI was performed in 160 subjects, from 2014 to 2018. These images were analyzed in the Radiology Department of the CCC and diagnosis and staging were made based on these images, histology and blood/urine tests. The data were de-identified and collected through patients' written informed consent and the Research Ethics Committee of the CCC approved the retrospective research study (annex I).

From the initial data set, 46 whole-body MRI were excluded due to coronal multiplanar reconstruction was not available or due to severe anatomical deformities (e.g. crooked spine). The general inclusion and exclusion criteria are listed as follows:

Inclusion Criteria

- Whole-body MRI was available;
- Coronal multiplanar reconstruction was available.

Exclusion Criteria

- Presence of implants (e.g. breast implants);
- Severe anatomical deformities (e.g. crooked spine);
- Severe distortion artifacts.

Thus, the final dataset included 112 WB-DWI images. Reminding the goal of this study and due to the anatomical differences, the dataset was divided according to gender: 49

women (44%) and 63 men, diagnosed with several neoplastic diseases: MM, Prostate Cancer (PC) or others[1]. The demographic details are listed in table 5.1.

Table 5.1: Demography of the female and male group.

Group	Age (average ± standard deviation)	Diagnosis		
		MM	PC	others[1]
Females	62 ± 9	49	0	0
Males	68 ± 10	46	11	6

Finally, patients were distributed in "Atlas construction"and "Validation", per gender, according to:

Inclusion Criteria

Male and female Atlas

- DWI images were available.

Male and female Validation

- DWI images (including low and high b-value images) and T1w images were available;
- Preference was given to those who had repeated images through time and to whom that had a big number of lesions.

Exclusion Criteria

Male and female Atlas

- Absence of an organ (e.g. only one kidney);
- Presence of any deformity in the hyperintense organs in DWI (e.g. presence of kidneys' cysts).

Male and female Validation

- Abrupt intensity difference between stations.

Based on the above criteria, they were grouped in: female atlas (32 female, 100% diagnosed with MM) and female validation (17 female, 100% diagnosed with MM), male atlas (42 male, 59,5% diagnosed with MM, 26,2% with Prostate Carcinoma (PC) and 14,3% others[1] and male validation (21 male, 100% diagnosed with MM). The demographic details of each group are listed in table 5.2.

[1]Follicular non-Hodgkin lymphoma, carcinoma of the transverse colon, nodular lymphoma, clear cell kidney carcinoma.

Table 5.2: Demography of the male and female atlases and male and female validation.

Group	Age (average ± standard deviation)	Diagnosis		
		MM	PC	others
Male atlas	69 ± 10	25	11	6
Female atlas	62 ± 10	32	0	0
Male validation	75 ± 11	21	0	0
Female validation	65 ± 9	17	0	0

5.2 Imaging Protocol

The data were acquired using a 1.5T Ingenia Philips scanner operated at the Radiology Department CCC. At least three MRI modalities were acquired for each patient: T1w, Short Tau Inversion Recovery (STIR) and DWI. In some cases, T2w images were also obtained.

Given that the main aim of this thesis is to remove the normal hyperintense organs from WB-DWI to further segment lesions and given the medical rationale to decide if a hyperintense region in DWI is or not a lesion (section 4.2.1), DWI and T1w images were used.

Four to five axial DWI sequences were acquired in different anatomical levels covering the WB of the patient, with 43-88 slices, using free breathing SS-EPI. Multiplanar reconstructions were made possible due to nearly isotropic resolution of the axial DWI datasets, in order to generate a single WB diffusion dataset. Each DWI sequence was acquired with 2 b-values (0 and 800 or 1000 s/mm^2). The high b-value images were used for further analysis, since malignant lesions in their majority are maintaining high signal intensity on high b-value images, due to restricted diffusion patterns. Table 5.3 and 5.4 summarizes the main parameters of DWI and T1w acquisition, respectively. Some imaging parameters are within a range since the imaging protocol has suffered some changes during the time the images were acquired (between 2014 and 2018).

Table 5.3: DWI acquisition parameters.

Parameter	Acquisition Value
TR	6219-12075 ms
TE	63-90 ms
Acquisition matrix	1628x902
Non-zero b-value	800 or 1000 s/mm^2
Slice thickness	4-7 mm
Pixel spacing	1.5-1.8 mm

Table 5.4: T1w acquisition parameters.

Parameter	Acquisition Value
TR	412 ms
TE	4 ms
Acquisition matrix	484x1219 or 484x980
Slice thickness	5-6 mm
Pixel spacing	2.09x2.09 or 1.14x1.14

5.3 Image Processing Steps

Figure 5.1 shows a representative coronal slice of a WB-T1w (left) and WB-DWI (right), used in this thesis. Organs as the spleen and kidneys show an hypointense signal on T1w, since their tissues have a long T1 (longer than TR, as explained in section 3.1.1). Due to high tissue density and cellularity, the brain, spinal medulla, spleen, kidneys, bladder and testis appear hyperintense on the high b-value DWI image (right).

Figure 5.1 also shows field in-homogeneities, especially between the first-second and fourth-fifth stations of the T1w image (left). Without pre-processing, this can result in misinterpretation of the images. On the T1w image, the leg's muscles are clearly the same tissue between the fourth and fifth stations, but their intensity is different. Plus, when radiologists read these images and look for possible MM lesions on the skeleton or extramedullary sites, the criterion that lesions have their intensity equal to or lower than that of muscle or nondegenerated intervertebral disks may lead to deceit. Also, some regions of the arms in the DWI image (right) show hyperintensities, due to the use of surface coils during image acquisition.

As an initial step, images were imported to Horos (*https* : *//horosproject.org*), where a visual inspection of potential artifacts, mismatch between different slices, missing slices or lack of signal were made. This also allowed to verify the acquisition parameters. Then, DICOM images were converted to Neuroimaging Informatics Technology Initiative (NIfTI) using an in-house built plug-in for Horos, which allows storage of matrices that encode alignment, registration and normalization results.

The techniques of registration, visualization and manual segmentation were implemented in Python language, using Pycharm as the interpreter and using the following free open-source libraries: SimpleITK [92], Simple Elastix [93] for registration, ITK--SNAP 3.6.0 (*www.itksnap.org*) [94] for manual segmentation and 3D Slicer 4.8.1 (*https* : *//www.slicer.org*) [95] for visualization and to test some SimpleITK filters. The imaging processing techniques comprised the following steps:

- Creation of a WB atlas per gender, based on the population available for study (42 male and 32 female);
- Manual segmentation of the normally hyperintense organs of each of the WB atlases, by four radiologists, on ITK-SNAP;

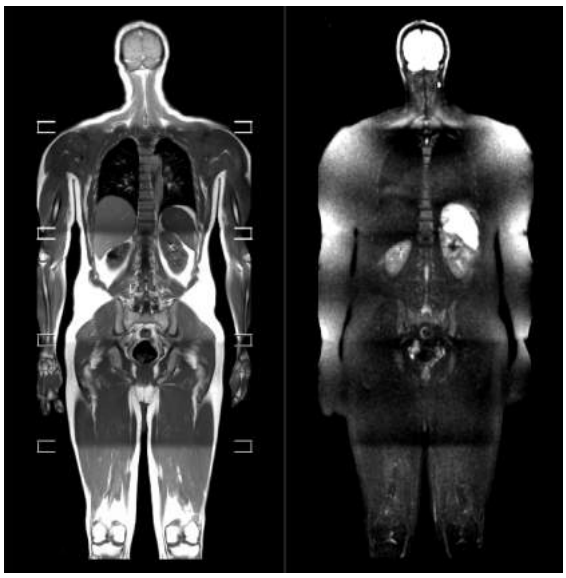


Figure 5.1: Representative coronal slice of a WB-T1w (left) and WB-DWI (right) of the same patient. The hypointense appearance of some organs as the spleen and kidneys on the T1w image is due to long T1. The hyperintense appearance of some organs on the DWI image is due to restrictive patterns on high cellularity tissues.

- Atlas accuracy assessment: evaluate if the registered hyperintense organs of the atlas overlap with a new DWI image;
- Application of the lesion detection algorithm to each validation image. This algorithm detects the normally hyperintense organs and possible lesions;
- Removal of the normally hyperintense organs, based on the manual segmented label;
- Statistical analysis. This step compares the semi-automatic segmentation with the manual one of four radiologists, considered the gold standard.

5.3.1 Atlas creation

The inter-subject registration/normalization with a template used to align the patient images and create a mean image of the population, atlas or template, plays a prominent role to remove the normal hyperintense organs in DWI images. This template image should represent, as much as possible, the population in study, more precisely, the position of the organs to be removed: brain, spleen, kidneys, testis and spinal medulla.

Taking into account the heterogeneity of the dataset, significant different anatomies, shapes and sizes were found in the WB images. For example, the organs' size and the distance between structures is never the same between different subjects. Consequently, as discussed in section 2.1.1, the literature predicts that deformable transformations should be employed for this kind of medical image registration, enabling flexibility of the structures. In the case of applying a local deformable transformation (B-spline), global

misregistrations must be previously corrected to avoid erroneous solutions, for instance using an affine transformation.

To fulfill this, an approach was developed and reformulated, which can be divided into single registration mean image creation and iterative mean image creation.

5.3.1.1 First Approach: single registration mean image creation

Firstly, a set of DWI images were selected from each male and female atlas group. One of the images, for each group, was selected as the reference and the other were registered to it. The reference image chosen had a normal size and shape. This algorithm was first tested with a small number of DWI images (n=9) and further increased to 32 female and 42 male DWI images, since the higher the number of similar WB-DWI images, the better represented the population is. Figure 5.2 shows eight moving images used for the first test, to be registered to one fixed image.

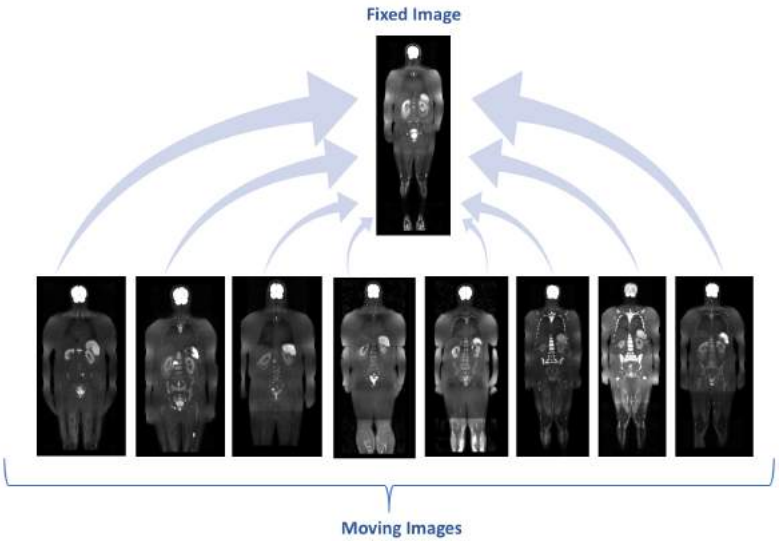


Figure 5.2: First registration test. Eight moving images from different patients were selected to be registered to a fixed image. Some of the images were just acquired up to the knees, while others went to the feet. Note also the different sizes and shapes of the different bodies and organs.

This first registration algorithm comprised a pre-registration and a more robust step, before applying a free-form deformation model with a regular grid of controlling points. Simple Elastix was used for the registration. The pre-registration was the computation of an optimal rigid transformation defined by three translational and three rotational parameters. This step was fundamental in order to get all the images in the same physical space, prior to a fine registration. This pre-registration step involved a multi-resolution registration with four levels using mutual information as the metric and the transformation was initialized based on the image's geometric center. The optimizer used was

an adaptive stochastic gradient descent with a maximum number of iterations of 255, a linear interpolator and a linear resample interpolator, which eased the process.

The geometric transformation found in the pre-registration step was then used as the initial solution for the optimization algorithm in the following step, which was an affine transformation. The transformation found in the affine transformation was then used as the initial solution to the free-form deformation model.

The affine transformation included twelve parameters, nine for rotation, scaling and shears and three for the shift, thus reinforcing the robustness of the alignment. The parameters were the same as the pre-registration.

For the free-form deformation model, a grid of controlling points was placed over the fixed image and the deformation field involved was determined based on the displacements of the grid. The size and voxel spacing of the fixed image was determinant to define the grid of controlling points. For example, considering the fixed image's size (1628 x 902 x 49) and voxel spacing (1.894 mm x 1.894 mm x 6 mm) used for the male atlas composed by 42 images, a realistic and accurate proposal was to set a grid point in every 150, 90 and 6 mm, in the x, y and z axis, respectively. The higher the number of grid points, the greater the computational time and the higher the probability of causing deformities to the image. The transformation between control points was chosen to be propagated by cubic B-spline, since they provide very good alignments with a low computational cost, as discussed in section 2.1.1. Because this was a final step, following up a good pre-registration, only one resolution was needed, which decreased the computational time.

The process was repeated until the stop criterion was achieved (for more detail, refer to [96]). To improve the registration and to discard false alignments, a mask of the WB of the fixed image (removing the background) was built and then morphologically dilated using a sphere with 10 voxels radius. After applying this registration algorithm to each moving image, the result were several images registered to the reference, all in the same physical space. Since the registered images and the fixed image had different intensity ranges, they were normalized by setting its mean to zero and variance to one. Finally, the normalized registered images and the normalized fixed image were added and divided by the number of images. To sum up, the final result was a mean image for each gender.

5.3.1.2 Second Approach: iterative mean image creation

The reference image chosen in the first approach was a DWI image with a normal size and shape. Still, the result was highly dominated by that reference image chosen. The literature suggests using, as the reference, an image as similar as possible to the rest of the population, in order to avoid biasing the results. Reminding this, the mean image previously built in the first approach was selected as the new reference (fixed) image and all the images were registered to it, using the previous transformations (first approach). Once again, the registered images were normalized, to adjust the intensity. Then, the

registered normalized images were averaged. This process was repeated three times, so as not to observe significant differences between the successive mean images. Figure 5.3 sums up the steps of the atlas (template) algorithm.

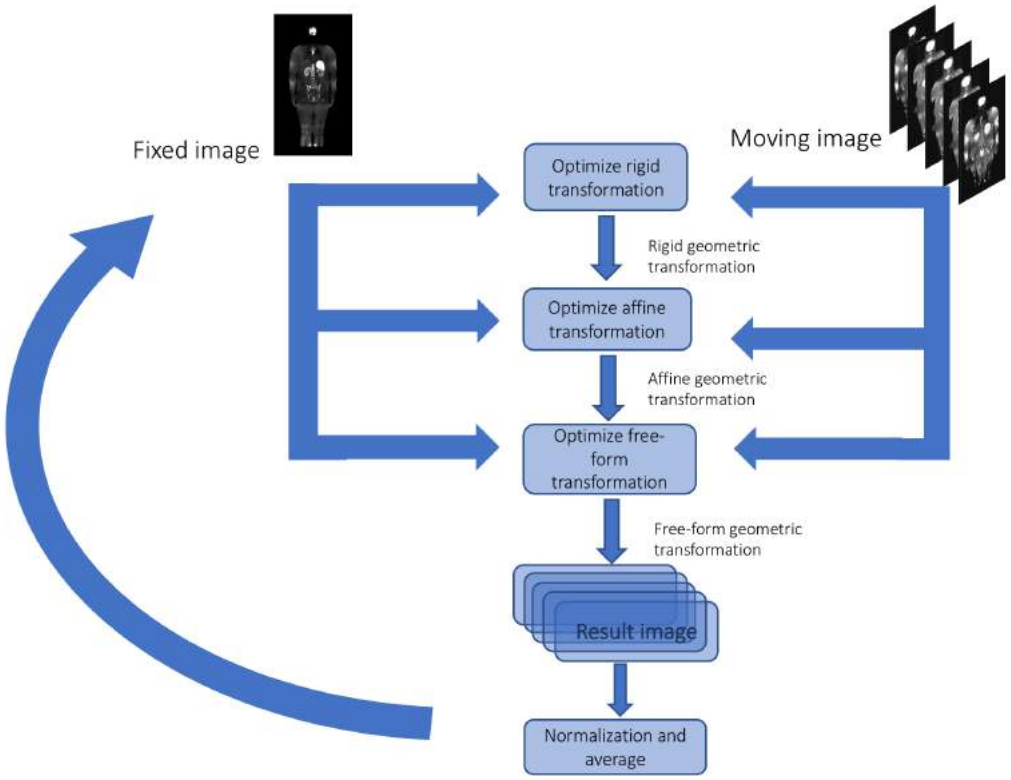


Figure 5.3: Atlas building scheme. First an image is selected as reference (fixed) and the rest are moving images. Each moving image is registered to the fixed image, by applying an optimal rigid transformation, followed by an optimal affine and free-form transformations (B-spline), which maximizes the similarity between the fixed and moving images. For each moving image registered, one image is obtained. Then, they are normalized, and an average is applied, resulting in a mean image. Then, this mean image is selected as the new fixed image and all the original moving images are again registered to it, following the same process. This was repeated until no meaningful changes were found between successive mean images.

5.3.2 Smart Semi-automatic lesion detection in DWI images

5.3.2.1 Manual hyperintense organs segmentation

Once the male and female atlas were built and the normal position of organs was known, the next step was to remove the hyperintense organs from the validation images, in order to distinguish lesions from normally hyperintense organs. To do so, four radiologists (3 specialists and 1 resident) manually segmented the brain, spleen, kidneys, bladder and spinal medulla in the female and male atlas, plus the testis in the latter one, using different labels. ITK-SNAP 3.6.0 was used for the segmentation. For now, these four

radiologists will be called experts (E). This was the only step that required an expert intervention and so it is considered a semi-automatic approach.

5.3.2.2 Lesion and hyperintense organs detection algorithm

As discussed in section 3.2, the literature suggests computing high b-value images (>1000 s/mm^2), instead of acquiring them, to achieve a higher contrast differentiation between normal and diseased tissues with a good SNR. Having that in mind, cDWI was obtained for each image in the male and female validation group, using equation 3.7. S_i was the image acquired with the higher b-value (800 or 1000 s/mm^2) and S_0 the lower, while b_i and b_0 were the b-values itself. Once the ADC parametric map was calculated, it was used to extrapolate the estimated signal intensity for each image voxel to any computed b-value. Literature recommendation of computing b-values of 1500 and 2000 s/mm^2 was followed. Therefore, at this point, there were available three b-values to validate the algorithm: one directly acquired (800 or 1000 s/mm^2) and two computed (1500 and 2000 s/mm^2). Besides the normal hyperintense organs in DWI, present in all three b-value, some of these images also showed hyperintensities in the arms, due to the use of surface coils during image acquisition.

Afterwards, an algorithm was applied to segment possible lesions and normal hyperintense organs in the DWI images, for the three available b-values. The algorithm chosen should consider the voxels' intensities and choose those who stand out, i.e., the higher intensities' ones. This could be done by setting a numerical value (threshold-based) above which would segment. However, since signal intensity in MRI depends on several factors (section 3.1), there is no single numerical value which would fit to all images to validate. Reminding chapter 2.2, several region growing algorithms were tested. However, since they relied on a list of seed points to initialize the method, this increased the validation and computation time, as well as introduced another user-intervention point, which was not the goal. To overcome this, an outlier removal approach (section 2.2.1.1) was used to define the optimal numerical value for each image, above which segmentation occurs. Since lesions show an hyperintense behaviour on DWI images, the upper outlier range was applied. This outlier value was different for each image, since the algorithm computed the first (Q1) and third quartile (Q3) per image. As suggested by [42], k values of 1,5 or 3 are likely to achieve good results. However, this suggestion is a generic statement about outliers, which was not done specifically to medical images. These recommendations were followed; however based on our preliminary tests, the k value was extended up to 4, since the suggested k-values were not valid to some WB images. This was applied to each image of the validation group, using the three different b-values available.

5.3.2.3 Removal of the normal hyperintense organs

At this point, there were available: one manually segmented label for each gender, that comprised the normally hyperintense organs of the templates; 21 male automatic outlier

removal segmentation and 17 female automatic outlier removal segmentation, for each of the three b-values. The reader should note that these automatic segmentation comprise not only possible lesions, as well as the normal hyperintense organs.

The next step was to remove these organs by registering the template segmented label to the same physical space as each validation image and to remove this registered label from the automatic outlier removal segmentation image. This registration could be performed by two different ways: inverse or direct. In the first way, the same transformation used to register each moving image to the atlas (section 5.3.1.1) is inverted and applied to the template segmented label, whose result should be a manually segmented label in the same physical space as the new validation image. However, since B-splines are not directly invertible and Simple Elastix does not provide a way to invert the transformation, this approach was not used. Instead, the direct approach was applied. Here, the atlas (moving) was registered to the new validation image (fixed) using the same methodology as in section 5.3.1.1. Then, the same transformation was applied to the template segmented label, whose final physical space was the same as each image to validate. Finally, each registered segmented label was subtracted from the correspondent automatic outlier segmented image. The final result were 21 male and 17 female binary images, where 1 corresponded to a possible lesion on DWI. For further analysis, these images were designated as Final Result DWI.

Figure 5.4 shows the semi-automatic lesion detection of DWI images scheme, which summarizes this section.

5.3.3 Automatic correspondence to T1w: more accurate lesion detection

As discussed in chapter 4.2 particularly for MM, the medical rationale for lesion detection does not only include visual inspection in DWI images. Thus, in order to achieve the gold standard's segmentation, T1w images were also used to improve lesion detection. The goal was to align each T1w image with each final result obtained in 5.3.2.3 (Final Result DWI) and decide if the lesion detected in DWI was coherent with the voxel intensities in T1w.

Firstly, due to the field in-homogeneities discussed in chapter 3, each T1w image had its intensity corrected, using the N4ITK MRI Bias Correction, available on 3D Slicer. Next, each corrected T1w image was registered to the same physical space as each Final Result DWI. Since the lower b-value DWI images (0 s/mm^2) has more anatomical information and their physical space is the same as the higher b-value images, and so the same as the Final Result DWI, lower b-value images were used as the reference and corrected T1w were used as the moving images. Given that this was an intra-subject, multi-weighting registration, it was not required to use a curved deformation model, since the images belong to the same subject and were acquired during the same examination without patient reposition with-in the MRI. Instead, literature suggests that a rigid registration can fit these images. However, it was important to perform a pre-registration step to

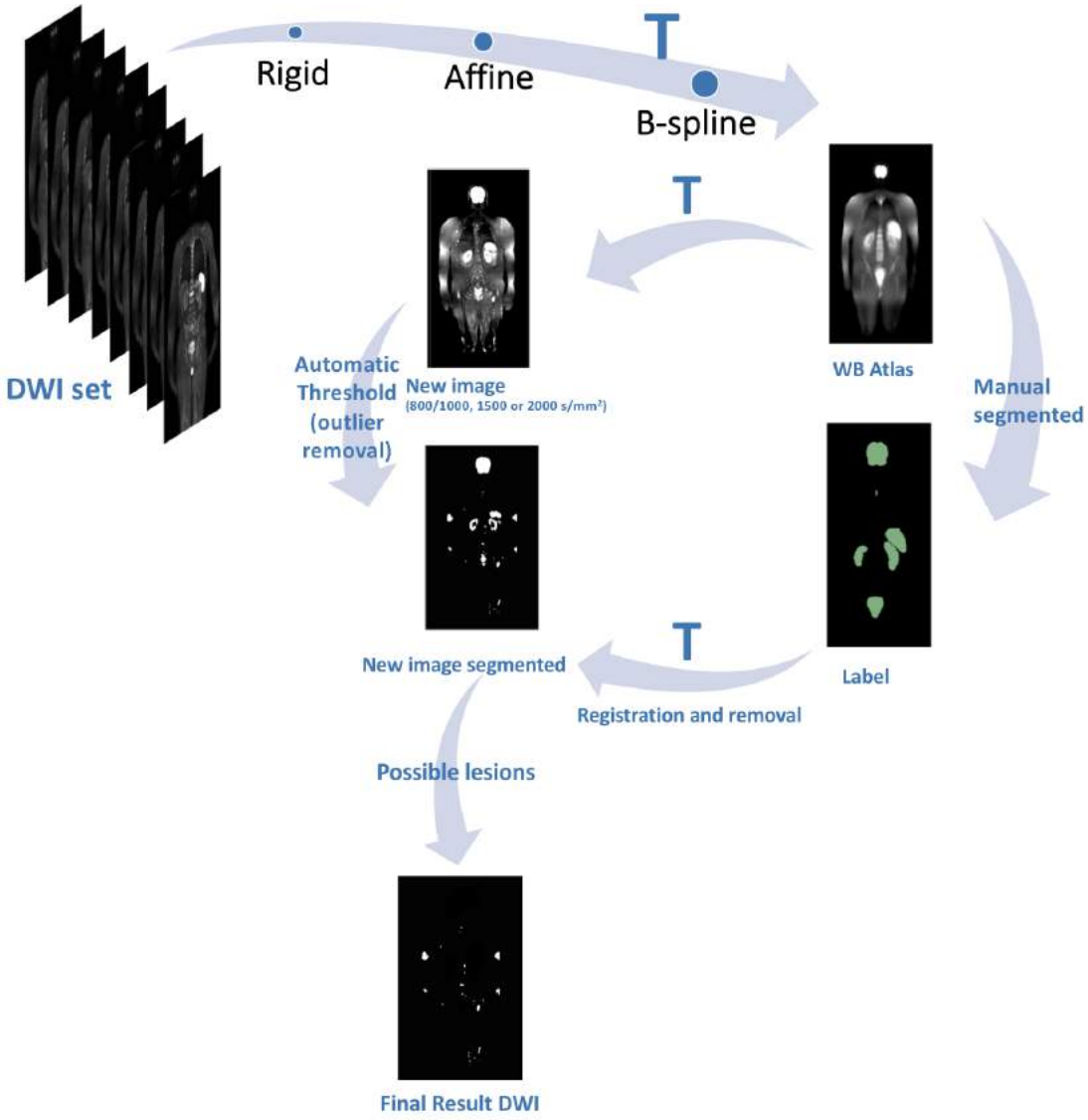


Figure 5.4: Semi-automatic lesion detection in DWI scheme. The WB atlas is built based on the registration of a set of DWI images, using optimized rigid, affine and B-spline transformations. The atlas is manually segmented by four radiologists, and their labels are merged (Label). Then, the atlas is registered to a new DWI image, not used to build the atlas, using the same transformation as before (T). Afterwards, an automatic threshold is applied to the new image, enhancing the normal hyperintense organs and possible lesions (New image segmented). Finally, the Label is registered to the new image segmented, using the same transformation (T), so that the normal hyperintense organs can be removed (Final Result DWI).

set the corrected T1w images to the same general registration space prior to a higher resolution registration. The pre-registration step was a translation alignment, followed by a rigid transformation, using a multi-resolution registration with four levels using mutual information as the metric. The number of spatial samples and the maximum number of iterations for each resolution were both increased to 50000 and 512, respectively, in order to achieve the best results. A third-order B-spline was used as the interpolator. Although this increased the computation time, this step was crucial to determine the exact location of lesions in T1w, making it necessary to implement those changes.

After having the corrected T1w images in the same physical space as the Final Result DWI, two approaches were developed to distinguish lesions from normal tissues: the voxel-by-voxel and the connected-component approach. Both approaches relied on the medical rationale that a MM lesion should appear hyperintense on diffusion but equal or darker than the muscle on T1w, as was described in section 4.2.1. Psoas was the chosen muscle to compare lesion's intensities, since it is normally a low-fat muscle and its area extends widely in more than one slice. A slice of the Psoas muscle (figure 5.5) was segmented from each of the corrected T1w images, and then registered to the same physical space as the Final Result DWI, using the same transformation as before. Then, the average of the Psoas muscle's label was calculated.

Figure 5.5: Example of a manually segmented label (blue) of a coronal slice of the Psoas muscle. The average of the intensity of this slice was used as reference to compare with possible lesions, seen on T1w images. This was done for every validation image.

Once the average intensity of the Psoas muscle was known, the first approach was to apply a threshold to the corrected registered T1w image, using the average intensity as the threshold. At this point, every pixel in T1w under this threshold value would be segmented. Then, this segmentation was intersected by a voxel-by-voxel approach with the Final Result DWI, resulting in a binary label whose pixels' intensity are lower

than the Psoas muscle in T1w but show high intensity in DWI, according to the outlier removal. However, this approach could discard some voxels since lesions are not always homogeneous.

The second approach grouped the Final Result DWI by connected labels, which means that if a voxel was connected with other voxel by a single point, they would be clustered in the same group. Then, instead of considering each voxel as an individual unit, the algorithm considered each connected group as a unique object. The goal was to have a label per lesion region, so that each lesion could be handled individually. Then, each connected label was overlapped with the corrected registered T1w and statistics were extracted (average, minimum of the label and percentiles), in order to decide either to eliminate or keep the label. In this case, if a certain percentile (between 1 to 100) of the label was less than the Psoas muscle's average, than it was considered a lesion. The higher the percentile used, the more probable the region is a real lesion, since its overall intensity is closer to be lower than the muscle's average.

The semi-automatic lesion detector algorithm took on average 10 minutes per patient, using a 3.1 GHz Intel Core i5 processor, with 8 GB of RAM, running macOS High Sierra.

5.4 Statistical analysis

As a final result of the image processing step, 17 female and 21 male segmentation binary labels were computed, per each of the three b-values (800 or 1000 s/mm^2, 1500 s/mm^2, 2000 s/mm^2), per T1w-approach (voxel-by-voxel or connected component approach), making a total of 102 female and 138 male labels to analyze. The reader should remind that the goal was to remove the normally hyperintense organs, using the atlas. As an extra step, it was also intended to find a label that segments lesions (if they exist) and does not include arm hyperintensities, referred in section 5.3.2.2.

To validate the AB algorithm, a radiologist and a researcher segmented the normal hyperintense organs of 5 males and 5 females DWI images, using 3D Slicer. The AB algorithm was applied to the same images and the organs' labels were compared. Three metrics (one overlap and two surface distance based measurements) were used to assess the agreement between the AB and manual segmentation. As discussed in section 2.2, DSC quantifies the overlap between two segmentation (1= complete match, 0 = no match). Hausdorff Distance (HD) or maximum surface distance is the maximal distance from a point in a set to the nearest point of another set. The smaller the distance, the better aligned are the sets (in this case, the segmentation). Average Hausdorff Distance (AHD) is the average of the differences between the nearest points of two sets. The last two metrics are surface distance measurements, which are expressed in millimeters. The three metrics were implemented on Python, using SimpleITK.

Finally, to validate smart semi-automatic lesion detector algorithm, the results were compared with each of the manual segmentation, done by four radiologists using 3D Silcer. Radiologists had access to T1w, T2w, STIR and DWI images. The manual lesion

segmentation were done within 6 weeks. Since the segmentation of the radiologist is considered to be the most accurate, variations from that are reported as erroneous. Similarity was quantified using the DSC. Positive Predictive Value (PPV), Negative Predictive Value (NPV) and sensitivity were assessed between the semi-automatic and the manual segmentation. PPV is the quotient between the number of true positive voxels and the total number of positives detected by the algorithm, expressing the proportion of positive voxels that are true positive. NPV is the quotient between the number of true negative voxels and the total number of negatives detected by the algorithm and express the proportion of negative voxels that are true negatives. Sensitivity is the number of true positives voxels detected by the algorithm divided by the true positive voxels detected by the manual segmentation, expressing the proportion of true positives that are correctly identified as such. The closer to one, the more similar is the algorithm's segmentation and the manual one. Also, in order to assess inter-radiologist variation, DSC was also used to quantify the matching between different radiologists. The four metrics were implemented on Python, using SimpleITK.

Chapter 6

Results and Discussion

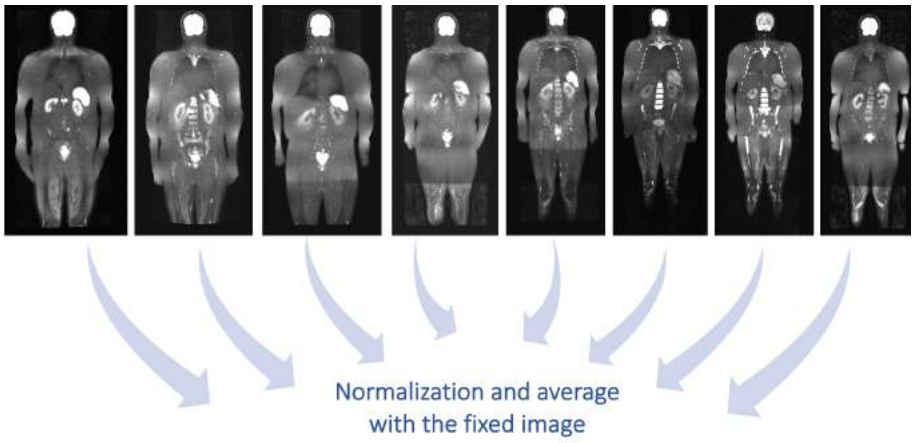
6.1 Atlas Creation

The following results are representative of the steps executed to create the male atlas. The female atlas was achieved in the same way and reached similar results, so only the final result and statistics will be shown.

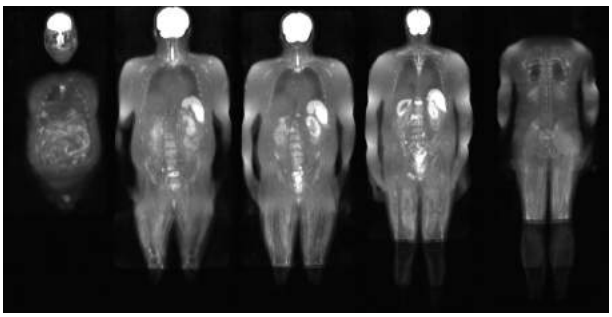
Figure 6.1 shows the coronal result of the registration and creation of the mean male image, using the nine images of figure 5.2 (section 5.3.1.1). Note that these images were built by applying an optimized rigid, affine and B-spline transformations. Then, the registered images and the fixed image were normalized and an average of all resulting images was obtained.

Although the transformations used were expected to achieve good alignments between WB images, the visual results show that the spleen, kidneys and vertebras are not yet completely overlapping. The forth slice shows a shadow in the left and right kidney, which means that there is, at least, one image that did not achieve a good registration. This could be explained by the fact that the fixed image chosen might not entirely represent the population of study and could be biasing the results. Plus, the fact that the kidneys and the spleen are showing good resolution, with a clear definition of their anatomy, confirms that this mean male image is highly influenced by the fixed image. The mean male image also shows two hyperintense areas, surrounding two vertebras. This is not critical, since the goal is to have a map of the normal hyperintense organ's location, which do not include the vertebras. Yet, ideally, this could be improved.

Due to pillow and position during acquisition and since the head is a mobile structure, some DWI images used as moving images showed the different positioning of the head on the sagittal plane. However, the brain appears to be well aligned 6.1. This could only be achieved by applying the pre-registration step before the high-resolution registration



a



b

Figure 6.1: First mean male image, built with nine DWI male images. a) The top eight images are the result of the rigid, affine and B-spline transformations, applied to the eight moving images of figure 5.2 section 5.3.1.1. Then, these eight images and the fixed image were normalized and averaged. b) The bottom image shows five representative coronal slices of the mean male image. Although the brain shows very good alignment, the kidneys, spleen and vertebras still need improvement.

(affine and B-spline transformations).

In order to succeed greater alignments between normal hyperintense organs, the fixed image should characterize all the images available. Figure 6.2 shows the second mean male image, obtained by using the previously built mean image as the reference and register the nine moving images (used to build the mean image) to it. The male template image's hyperintense organs show lower resolution than the mean male image's hyperintense organs, since the original fixed image contribution is decreasing. Thus, the spleen is reaching a more similar shape with rest of the population. The reason why the same is not happening to the kidneys and bladder is because they have lower intensity, thus will benefit from an increase in population. Plus, the bladder intensity depends on whether it is full or not. It is also noticeable that the hip bones are not overlapping. Again, this is not a problem since they are not hyperintense organs. The two hyperintense areas found previously on the mean male image are not seen in this male template, which confirms that using a mean image of the population as the fixed image normalizes the contribution

of each image. Once again, the brain shows very good alignments.

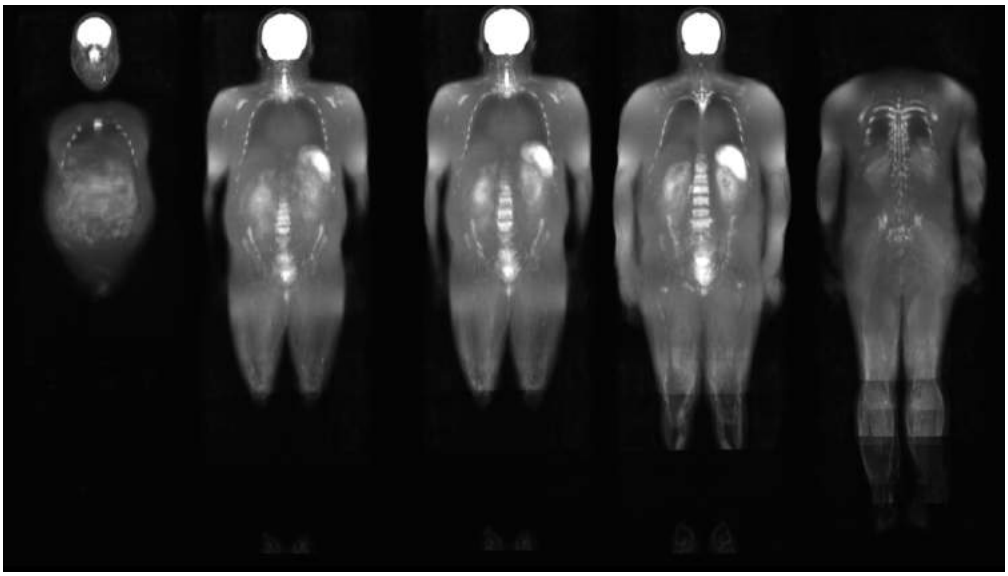


Figure 6.2: First male template image, built with nine DWI images and using the previously built mean image as the fixed image. These five representative coronal slices show, once again, very good alignments of the brain. The spleen shows a different shape, when compared with the mean male image, since the fixed image weighting is decreasing. Thus, the contribution of all available images is increasing, removing the bias of the fixed image.

The result could still be improved by using the newly built template as the new fixed image and repeat this until no noticeable differences are found between consecutive template images. Figure 6.3 shows the third mean male using the same nine moving images and the previous male template as the reference. Although the differences are less noticeable, it is evident that the kidneys, hip bones and vertebras are converging. No visible differences are found in the brain, spleen and bladder. This was the best result attained with nine images, so it was called template image.

Figure 6.4 shows the result of increasing the number of male DWI images to 18, using the same rationale as before. As expected, by increasing the number of images, the alignment of the normal hyperintense organs is better: the kidneys show a higher degree of overlapping, the location of the bladder is converging and the brain and the spleen show similar results. As was already stated, although the vertebras are not overlapping correctly, it is not problematic.

Seven coronal representative slices of the final male and female atlas are shown in figures 6.5 and 6.6, respectively.

These male and female atlases were built with 42 and 32 images, respectively. As expect, there are visible anatomical differences between them: the female atlas shape is rounder on the hip and chest areas, the female organs appear to be smaller and also the presence of the testis on the male atlas, which justifies the decision of building an atlas for each gender. The normal hyperintense organs (brain, spleen, kidneys, spinal medulla

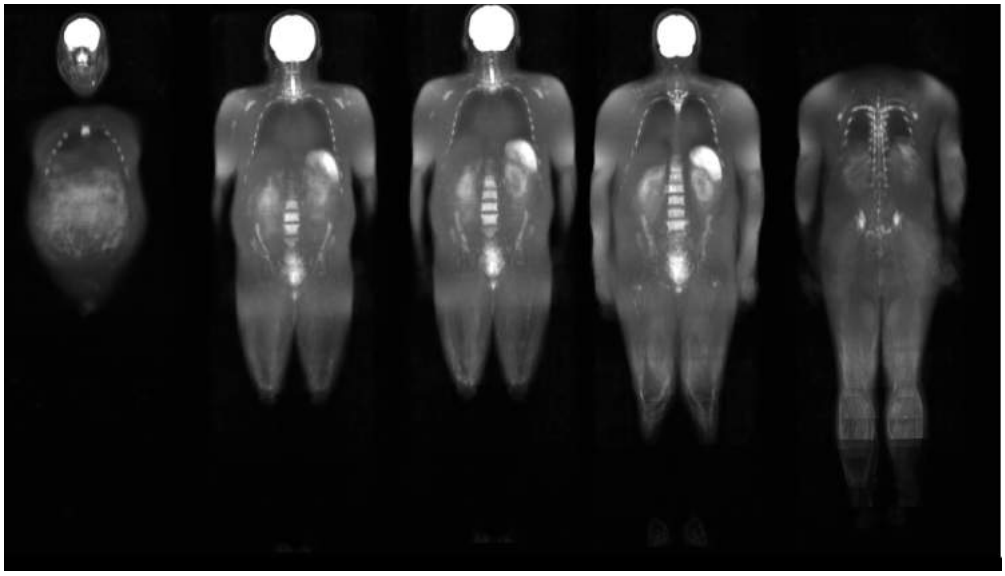


Figure 6.3: Male template image, built with nine DWI images and using the previously built template image as the fixed image. These five representative coronal slices do not show differences in the brain, spleen and bladder, when compared to the first male template. The kidneys, hip bones and vertebras are converging to specific areas, achieving better alignments.

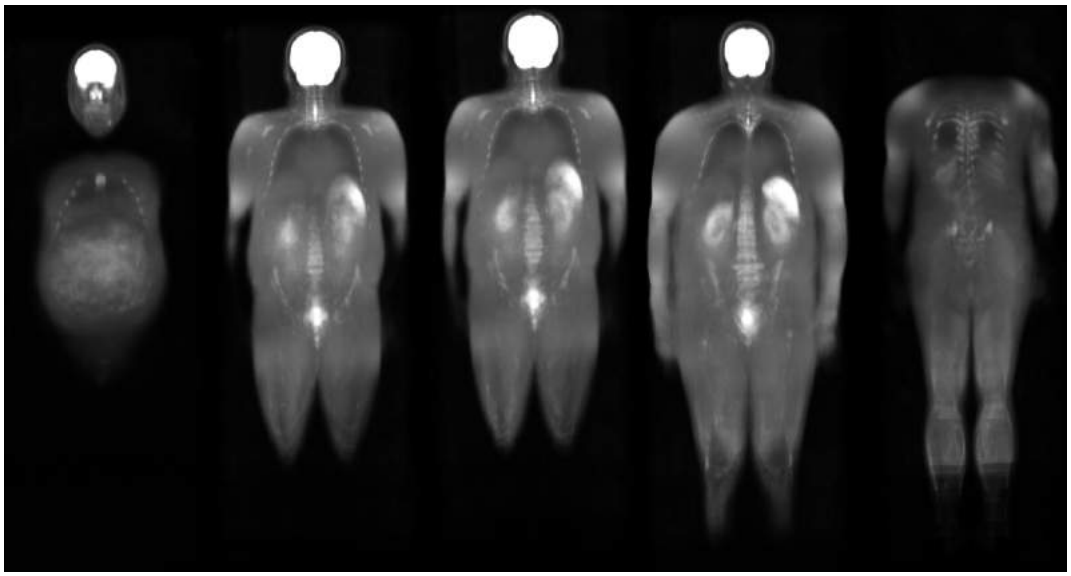


Figure 6.4: Male template image, built with 18 DWI images and using the first template image built with 18 images as the fixed image. These five representative coronal slices do not show differences in the brain, spleen and bladder, when compared to the first male template. The kidneys, hip bones and vertebras are converging to specific areas, achieving better alignments.

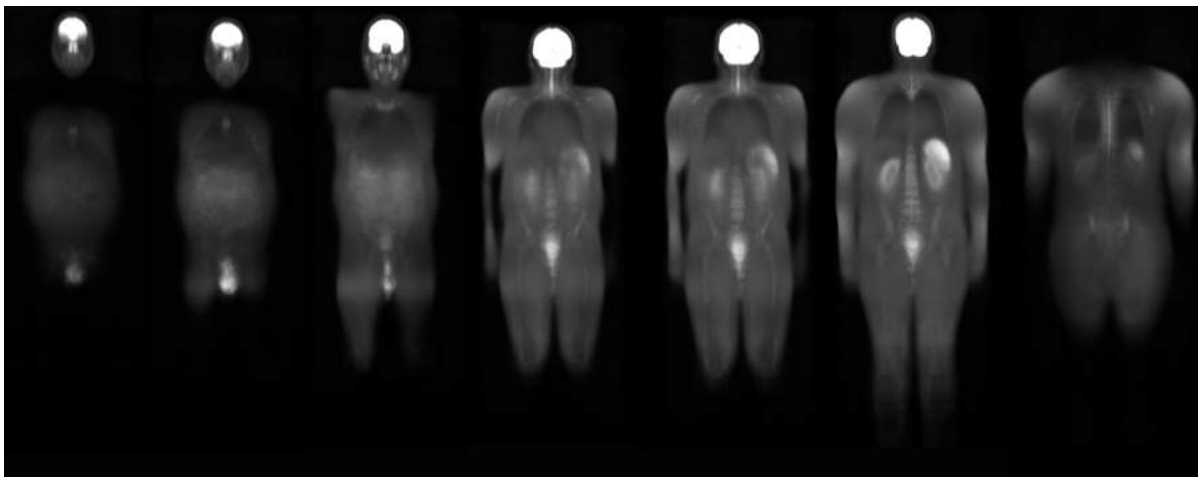


Figure 6.5: Final male atlas, built with 42 DWI male images. These seven coronal representative slices show that the normal hyperintense organs achieved good alignments, using a mean image as the fixed image.

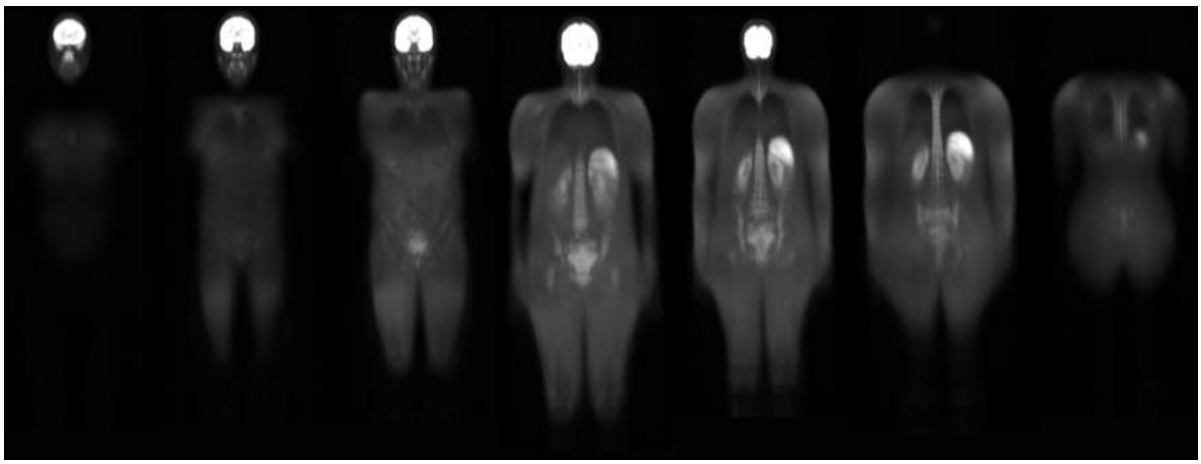
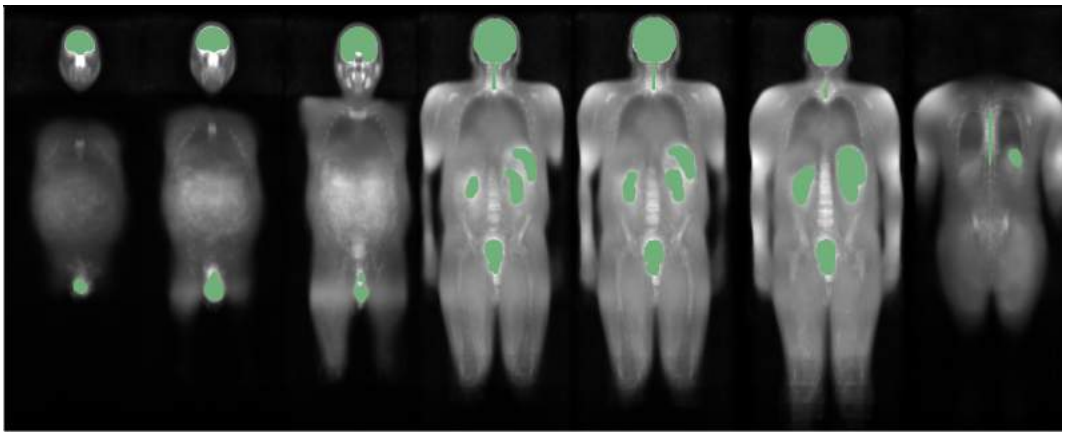


Figure 6.6: Final female atlas, built with 32 DWI male images. These seven coronal representative slices show that the normal hyperintense organs achieved good alignments, using a mean image as the fixed image.

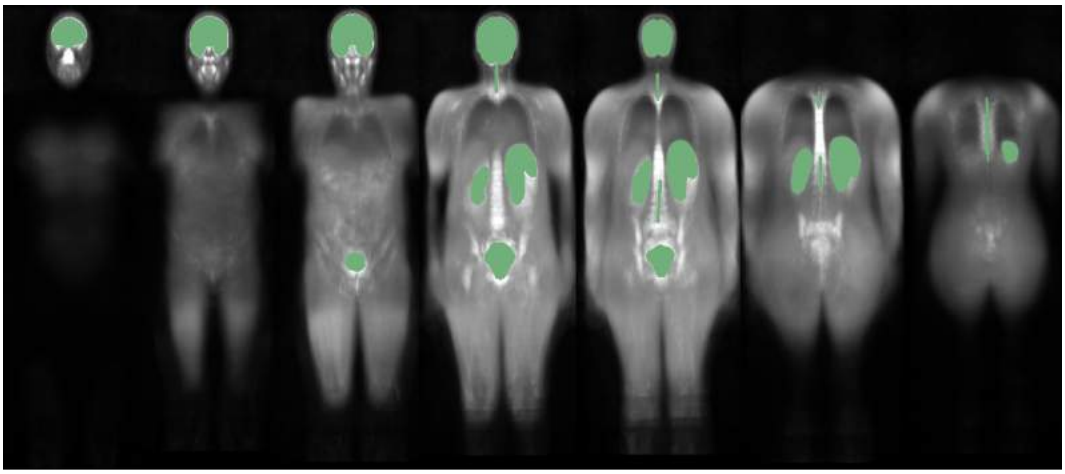
and testis) are well defined, thus its average position is well known. This corroborates the fact that the algorithm is optimized for a large number of whole-body DWI images, even with different shapes, sizes and gender. Despite being built with a smaller number of DWI images, the female atlas does not show significantly observable organ's overlapping differences with the male atlas.

The green label in figure 6.7 shows the union between the four radiologists' manual segmentation of the brain, spleen, bladder, spinal medulla, testis in the male and female atlas.

Both the male and female atlases algorithm can only be truly evaluated when a new DWI image (not used to build the atlas) is registered to the manually segmented labels. If the registered label overlaps with the normal hyperintense organs of a new DWI image,



a



b

Figure 6.7: Manually segmented hyperintense organs (green) of the final male atlas (a) and final female atlas (b). The four radiologists' labels were merged, in order to obtain a single male and female label.

the algorithm is successful. In order to perform this analysis, the normal hyperintense organs of five male and five female DWI images were segmented by a radiologist and by a researcher familiar with this type of images. These images were not used to build the atlas. The developed transformations were applied to the labelled hyperintense organs of the male and female atlas, which were deformed to fit each of the five male and female DWI images. Figure 6.8 shows the result of the semi-automatic segmentation using the previously built atlas, using as input a male DWI image. The brain and spinal medulla labels seem to be covering the correct area in this slice, while the other organs are over segmented, with greater emphasis on the bladder.

DSC, AHD and HD were extracted for each anatomical label (brain, right kidney, left kidney, spine, bladder, testis, spleen) between the manually segmented and the semi--automatic label. These metrics were also extracted for the junction of all labels, since the algorithm removes these organs as an unique label.

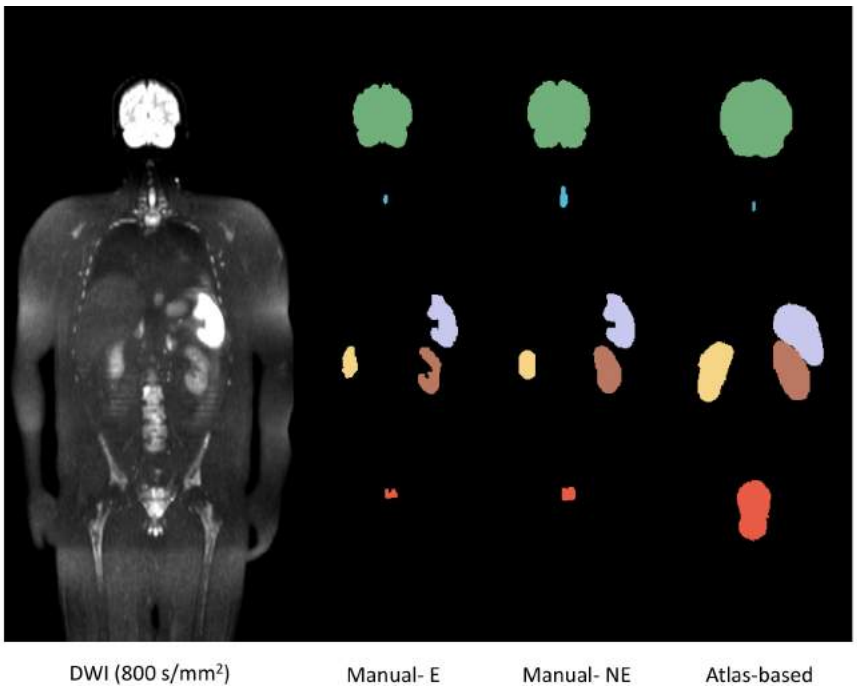


Figure 6.8: DWI representative coronal image, manual and automatic segmentation of hyperintense organs (brain- green, spinal medulla- blue, spleen- purple, left and right kidneys- brown and yellow, bladder- red and testis). One of the manual segmentation was performed by a radiologist (Expert- E) and the other was done by a researcher (Not-Expert NE). In this slice, the testis are not visible.

Tables 6.1 and 6.2 show the pooled mean metrics ± standard deviation from the segmented labels for the AB algorithm versus the radiologist (E) and the researcher (NE). These metrics were also extracted for the radiologist versus the researcher segmentation.

Table 6.1: DSC, AHD and HD ± standard deviation for each anatomical label, segmented manually by a radiologist (Expert- E), a researcher (Not-Expert NE) and Atlas-based (AB), when using male DWI images as input only.

		Brain	Right Kidney	Left Kidney	Spine	Bladder	Testis	Spleen	Complete Label
E vs AB	DSC	0.80±0.07	0.51±0.13	0.50±0.14	0.15±0.03	0.12±0.12	0.17±0.09	0.44±0.16	0.63±0.03
	AHD	1.39±0.73	5.34±3.72	4.74±3.10	10.77±1.83	29.84±19.15	13.66±2.34	7.76±5.00	5.41±1.63
	HD	231.17±256.59	44.99±17.79	33.32±11.14	65.09±24.53	95.40±25.77	59.42±5.85	55.26±19.64	87.6±16.95
NE vs AB	DSC	0.81±0.06	0.56±0.15	0.56±0.14	0.15±0.03	0.13±0.15	0.19±0.06	0.48±0.17	0.65±0.03
	AHD	1.35±0.69	5.26±4.23	4.30±3.23	11.06±1.74	29.16±18.19	13.39±1.18	7.09±5.16	5.13±1.63
	HD	29.31±5.53	46.02±21.77	31.49±11.64	61.93±15.28	95.92±25.17	57.78±5.33	52.48±19.9	83.72±13.61
E vs NE	DSC	0.97±0.01	0.87±0.04	0.86±0.07	0.65±0.05	0.87±0.06	0.81±0.05	0.86±0.05	0.93±0.01
	AHD	0.08±0.02	0.41±0.17	0.49±0.33	1.3±0.18	0.37±0.22	0.7±0.39	0.61±0.42	0.25±0.07
	HD	240.64±263.16	12.79±4.93	11.93±5.58	28.17±5.60	12.05±4.85	14.77±7.48	16.30±8.69	25.17±4.87

A DSC above 0.7 is suggestive of a good overlap [97]. Considering the complete label, it is seen that the agreement between manual segmentation gives the highest mean DSC (0.93±0.01 for male, 0.94±0.01 for female), the lowest AHD (0.25±0.07 mm for male, 0.15±0.05 mm for female) and the lowest HD (25.17±4.87 mm for male, 22.78±8.85 mm for female) than the AB algorithm either versus the specialist or the researcher. However,

Table 6.2: DSC, AHD and HD ± standard deviation for each anatomical label, segmented manually by a radiologist (Expert- E), a researcher (Not-Expert NE) and Atlas-based (AB), when using female DWI images as input only.

		Brain	Right Kidney	Left Kidney	Spine	Bladder	Spleen	Complete Label
E vs AB	DSC	0.77±0.09	0.31±0.12	0.47±0.15	0.07±0.06	0.11±0.04	0.53±0.09	0.58±0.05
	AHD	1.97±1.48	12.35±8.15	5.12±2.57	10.24±2.38	15.24±5.47	4.03±1.09	5.86±2.08
	HD	32.35±15.45	58.21±24.71	36.23±5.25	54.39±13.33	68.45±11.69	36.00±1.45	68.63±6.69
NE vs AB	DSC	0.78±0.09	0.33±0.13	0.51±0.15	0.08±0.06	0.12±0.05	0.55±0.09	0.60±0.06
	AHD	1.90±1.51	12.21±8.32	4.68±2.27	10.26±2.42	14.97±6.02	3.83±0.98	5.72±2.19
	HD	31.63±15.53	57.84±24.38	34.99±5.5	53.12±16.85	67.18±12.8	34.70±1.49	67.45±6.26
E vs NE	DSC	0.96±0.01	0.92±0.02	0.90±0.03	0.75±0.03	0.87±0.03	0.93±0.03	0.94±0.01
	AHD	0.09±0.02	0.22±0.09	0.34±0.13	0.75±0.11	0.35±0.17	0.20±0.12	0.15±0.05
	HD	10.66±2.64	11±7.22	11.25±3.29	20.46±10.65	7.87±3.22	8.47±2.23	22.78±8.85

it is important to note that the performance of the AB method is favorably comparable to both of manual segmentation. A relatively good mean DSC was returned (0.63±0.03 for male, 0.58±0.05 for female), a small AHD (5.41±1.63 mm for male and 5.86±2.08 mm for female) and a relative high HD (87.60±16.95 mm for male and 68.63±6.69 mm for female) when comparing the specialist and AB segmentation.

Keeping in mind the HD definition and given the pixel's dimensions of the images (~0.155x0.155x6 mm), it is clear that failing a segmentation by a pixel in the inferior-superior axis (z axis) will always result in an error greater than 6 mm (distance between two consecutive axial slices). For instance, the HD between the researcher and the AB algorithm for the male images of the brain (29.31±5.53 mm) and of the right kidney (31.49±11.64 mm) might be only failing by ~5 pixels on the inferior-superior axis.

Although not outperforming the manual segmentation results, the AB method achieved very good results for the registration of the brain, both for male and female. Since the head is a mobile structure and due to the pillow and position during acquisition, it was often found on the sagittal plan that the head and neck would be placed with different angles.The pre-registration step (rigid transformation) was the key to achieve a very good registration of the brain, since it provided an initial alignment before a thinner registration. Without this step, we verified that different heads did not overlap, even though the body overlapped almost completely. Right and left kidneys and spleen achieved medium results, because its position and size is highly variable. The spine, bladder and testis were the organs that achieved the worst results regarding overlap and surface distance measurements. The fact that the spine is very thin and may not be straight, makes the registration of this structure a challenging task. Similarly, the segmentation of the bladder depends upon it being full or not. When the radiologists segmented the atlas, they considered a full bladder (figure 6.7). When using that label to be registered to a DWI image with an empty or almost empty bladder, it is likely that the results are much worse, as seen on tables 6.1 and 6.2. However, this does not mean that the organ cannot be successfully removed. In fact, the bladder label resulting from the AB algorithm is generally over segmenting, surrounding the whole bladder area and excluding nearby structures,

as bones. Plus, even over-segmenting, if the bladder is not full, it will not be hyperintense on DWI images. Thus, it will not interfere with lesion detection. Likewise, the testis are mobile structures which, due to their position while the patient is lying on the MRI platform, may appear in different slices, which justifies the small agreement between the AB algorithm and the manual segmentation. Although hyperintense on DWI when full, the gall bladder was not included on the normal hyperintense organs since it does not have a fixed anatomical place. Also, it is not always full.

Given the size of the hyperintense organs, it is correct to affirm that these reported distances (HD and AHD) are not significantly relevant. Distances of the order of 30 mm to 10 cm are not critical to most of the organs. Deviations from this are reported by the HD measured from the brain of the male images, 231.17±256.59 mm and 240.64±263.16 mm, when comparing the specialist with the AB algorithm and the researcher, respectively. However, in this case, the DSC is very close to one.

Although being adequate to define and identify organs, manual segmentation of WB DWI images is a very slow, time consuming process. On average, manual contouring of each WB image took 20 minutes. Comparing to the 6 minutes that the AB algorithm took to segment the hyperintense organs of an image, our algorithm represents a solution for the time constriction.

To our knowledge, there is no previous relevant work describing a semi-automatic segmentation of normal hyperintense organs on WB-DWI, using an AB approach. Thus, the performance of our method cannot be directly compared to any method described on the literature. The most similar work that was found compared the performance of Classification Forests, Convolutional Neural Networks and Multi-Atlas approaches to segment several abdominal organs in WB-MRI, using DWI images as input for the Multi-Atlas approach [2]. Although Multi-Atlas is most likely to achieve better segmentation than simple Atlas-based approaches, since represents better the intersubject variability of the anatomy to be segmented (section 2.2.6), we decided to compare the results to have insight about other similar studies. The results of the Multi-Atlas study reported the following mean DSC: Right Kidney 0.77±0.07; Left Kidney 0.72±0.13; Spleen 0.58±0.14; Bladder 0.69±0.23. Comparing to our work, the DSC of the kidneys and spleen are similar to those of this study, reinforcing the difficulty to register these structures, possible due to variations of size and position. The bladder, on the other hand, has a significantly different DSC. Here, a Multi-Atlas approach seems to be better, adjusting to different bladder sizes. This difference could also be explained based on the differences in the registration algorithm used in this study.

Despite validating only 10 images (five per gender), this small validation test allowed to have a proof of concept that AB segmentation enables accurate semi-automatic removal of multi hyperintense organs in WB-DWI of a set of neoplastic patients with good agreement to the manual segmentation. By completing this semi-automatic multi hyperintense organ removal in WB-DWI, the main goal of this thesis was achieved. The success of this step was crucial to move on to the next step, which is the ultimate goal of the project

where this thesis is included: automatic identification and segmentation of lesions.

6.2 Smart Lesion Segmentation

The purpose of this algorithm is to assist radiologists when defining lesions and further assess tumor burden. This section is divided in the computation of the b-values and lesion detection. First, two b-values were computed and the optimal one was chosen to be used for further analysis, together with the directly acquired one. Then, the algorithm was applied to both b-values of the validation images. Semi-automatic segmentation results were compared with the manual ones of the radiologists.

6.2.1 Computed b-values

b-values of 1500 and 2000 s/mm^2 were computed on 21 male and 17 female DWI images. For reasons of content limitation, only one representative example per gender will be displayed.

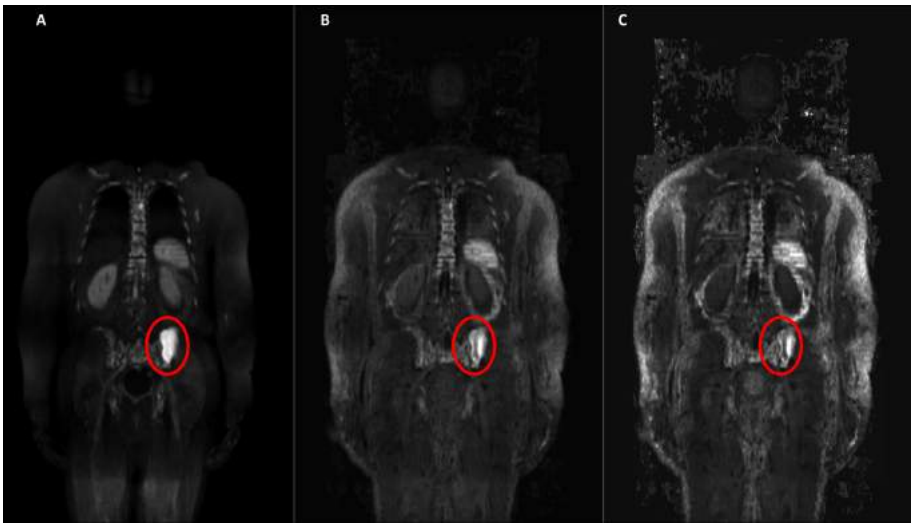
Figure 6.9 shows a side by side comparison between three different b-values: one directly acquired from MRI (A- 800 or 1000 s/mm^2), and two computed (B- 1500 s/mm^2 and C-2000 s/mm^2) from a male (a) and women (b). As expected, there are some structures that have their intensity decreased when using these computed b-values, enhancing diseased tissues over normal ones. The kidneys, that are hyperintense on the 800 s/mm^2 image, are a clear example of this. It is also possible to distinguish a lesion, hyperintense on the three images, represented by a red circle. However, the 1500 s/mm^2 b-value image displays a higher contrast differentiation between normal and disease tissues, when compared with the 2000 s/mm^2. Some regions of the arms show high intensities due to the use of surface-coils when acquiring the image. Although present in all images, it is most notorious when the b-value is increased. Similarly, the noise increases with the b-value. As a consequence, defining the lesion and differentiate it from other tissues using the 2000 s/mm^2 b-value was a very difficult task to be performed by a threshold algorithm, invaliding the use of this computed b-value.

From this point, the 2000 s/mm^2 b-value image was discarded from subsequent analysis.

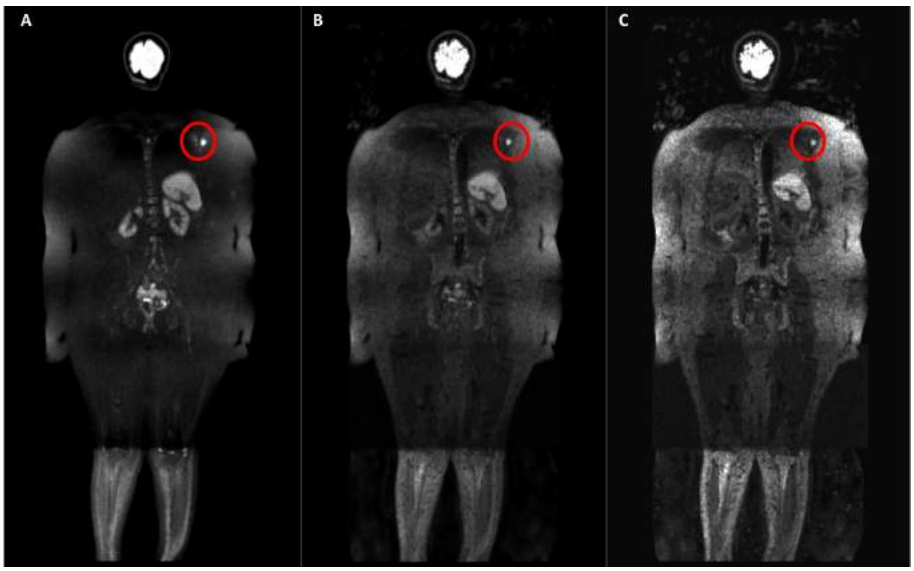
6.2.2 Lesion Detection

Lesion detection was performed on 21 male and 17 female DWI images. DSC, PPV, NPV and sensitivity between the smart and manual segmentation can be consulted on tables A.1, A.2, A.3, A.4, A.5, A.6, A.7, A.8 on appendix A. For reasons of content limitation, only the representative images per gender will be shown.

In order to the results of the manual segmentation can be coherently compared to the smart segmentation ones, it was defined that if at least one of the radiologists did not find a single lesion on the images, this case would not be used to assess the overall mean DSC,



a



b

Figure 6.9: Three different b-values applied to a male (a) and female image (b). A- b--value acquired directly from MRI (800 or 1000 s/mm^2); B- computed b-value (1500 s/mm^2); C- computed b-value (2500 s/mm^2). The red circle contours a possible lesion, hyperintense in all b-values. It is possible to see the contrast differentiation between normal and diseased (red circle) tissues when increasing the b-value. However, the 2000 s/mm^2 b-value image is very noisy, being hard to distinguish a lesion from the arm hyperintensities, due to the use of surface coils.

PPV, NPV and sensitivity of manual and smart segmentation. This is easily explained by the fact that the algorithm was created to find and delineate lesions. If there is a question whether a lesion is present or not or even if it is not present, then the algorithm should not be used.

Although the algorithm used was the same for all validation images, varying k and

the percentile used (see sections 5.3.2.2 and 5.3.3), each case is different. The lesions found are very heterogenous and each patient is different. Thus, the results of lesion detection will be presented and discussed per scenario, grouping images (cases) with particular features, which were important for this thesis explanation. Also, throughout this section, there will be presented seven sub-images per scenario. Sub-image A will display a representative coronal slice of DWI (b-value 800 or 1000 s/mm^2) and T1w, left and right respectively. Sub-images B and C will exhibit two different coronal slices manually segmented by four different radiologists in yellow (E1), purple (E2), red (E3) and dark green (E4). D and E will show different approaches of the smart algorithm, using the 800 or 1000 s/mm^2 b-value image. From left to right, we will be able to see the Final Result DWI (blue) and the integration with the T1w image, using the voxel-by-voxel (pink) and connected-component approach (light green), respectively. Similarly, F and G show the exact same coronal slices and same algorithm but applied to the 1500 s/mm^2 b-value image.

Male

- **Scenario 1**

  This scenario groups Case 1 and 2 of the same patient (one year apart). The pattern is considered focal with one big lesion on the iliac bone and other small areas on the skeleton.

  Figure 6.10 shows several comparisons between manual and smart segmentation of Case 1. From the manually segmented sub-images (B and C), we can verify that there is high agreement (mean DSC 0.772) on defining big lesions, while the small ones are often discarded by some radiologists. For instance, there is a possible lesion on the right humerus and iliac bone that are only segmented by 2 radiologists. The Final Result DWI using the 800 s/mm^2 b-value image (D and E) applied k=4, while k=3 was used to obtain the Final Result DWI for the 1500 s/mm^2 image. These k values were the optimal ones that covered all lesions segmented by the radiologists.

  It is clear that both blue labels (Final Result DWI using 800 and 1500 s/mm^2 b-values) are over segmenting. D shows the bladder wrongly segmented, because it is full, so hyperintense on 800 s/mm^2 DWI image. Increasing the b-value to 15000 s/mm^2 seems to improve this, since it enhances contrast differentiation between different tissues. However, as was discussed in the previous section, this high b-value image also increases the arm hyperintensities. Consequently, this contribution remains even when incorporating T1w information. The voxel-by-voxel approach (pink) is discarding some voxels, since the criteria is to segment when its intensity is lower than that of the average of the Psoas muscle. When the radiologist does this analysis, he does a visual inspection of what seems to be lower than the muscle, instead of a numerical analysis. Plus, lesions do not always have their intensity

values homogenous. Therefore, the voxel-by-voxel approach has its detection value decreased, when compared with the connected-component approach (light green). This last approach shows best result when applied to the 1500 s/mm^2 image, since it detects the whole big lesion area, does not detect the bladder and also shows coherence with the majority of the other lesions segmented by the radiologists (mean DSC 0.468). The percentile used was 39th, in order to cover all lesions segmented by the radiologists. Although the connected-component approach using the 800 s/mm^2 image is segmenting part of the bladder, this is not a problem for the radiologist, since MM lesions are mainly located in the skeleton or nearby. Exceptions are the extramedullary lesions which are less common [98]. Plus, the radiologist can erase what he believes is not a lesion. Although the mean DSC of the computed b-value image (0.468) is higher than the directly acquired image (0.411) when using the connected-component approach, its sensitivity is lower (0.80 vs 0.73). While the first is detecting the arm hyperintensities, the second is detecting part of the bladder (smaller area), hence the difference.

As can be consulted on table A.3, the difference between the mean DSC of the smart segmentation tool (0.468) and the manual segmentation (0.772) is explained by the high mean NPV (1.000), due to the segmentation arms in the automatic tool. This means that the smart algorithm is detecting extra areas that were not pointed by the radiologists because they were obviously not MM lesions. More importantly, the algorithm is not failing to detect the areas that were segmented by the radiologists as lesions, stressing its potential (substantially high PPV). In fact, the over segmentation of the tool can be easily corrected by the radiologists when using this semi-automatic smart tool, and yet with advantages in terms of time, precision and reproducibility. By helping the radiologists reading exams, the examination time is decreased and systematic and the false negative rate of detecting lesions is significantly lowered.

The smart segmentation results of Case 2 were very similar to that of Case 1, since they were images from the same patient with a one-year interval. In this case, the connected component approach applied to the 800 s/mm^2 showed the best results, because the arm hyperintensities' region was larger on the computed b-value.

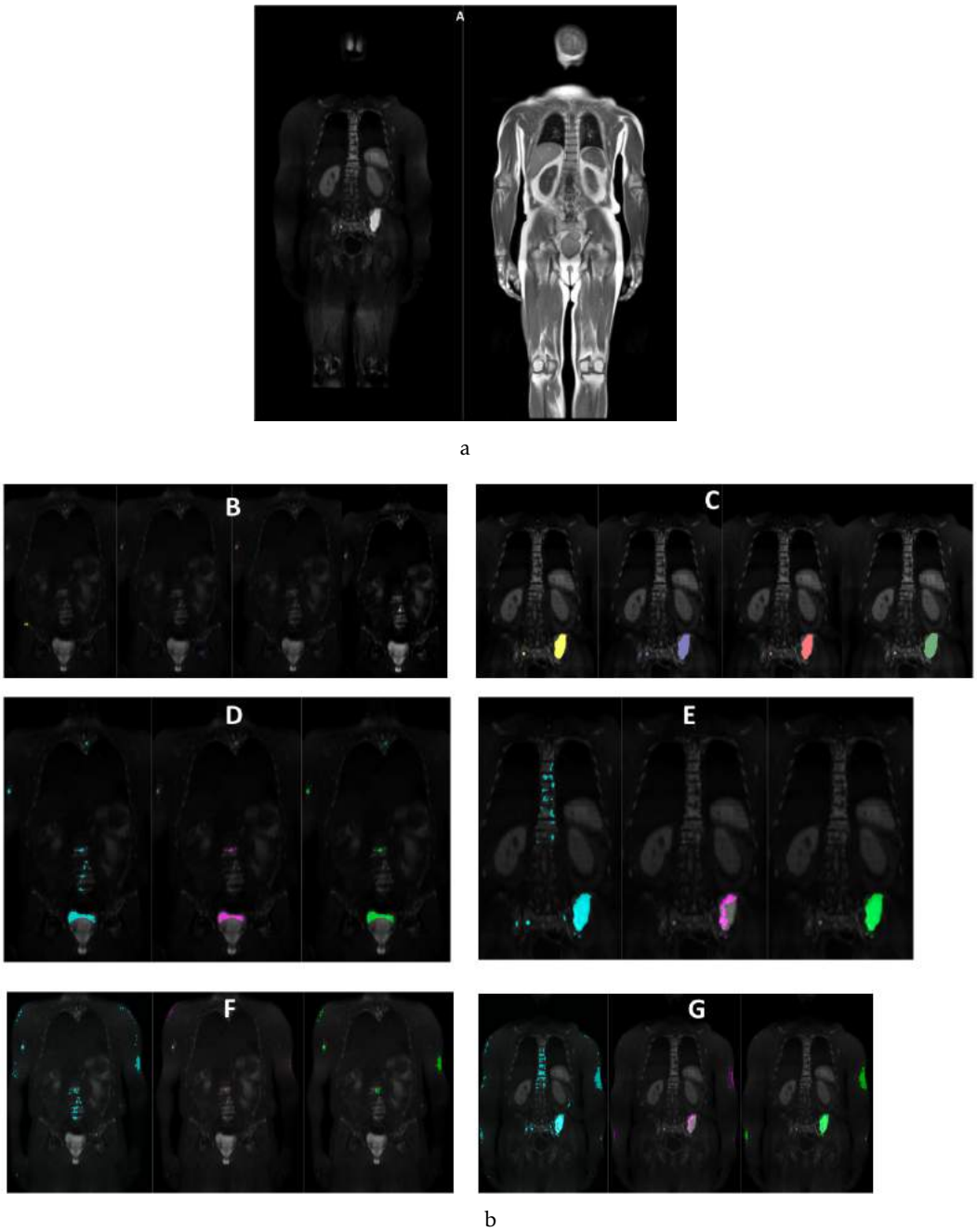


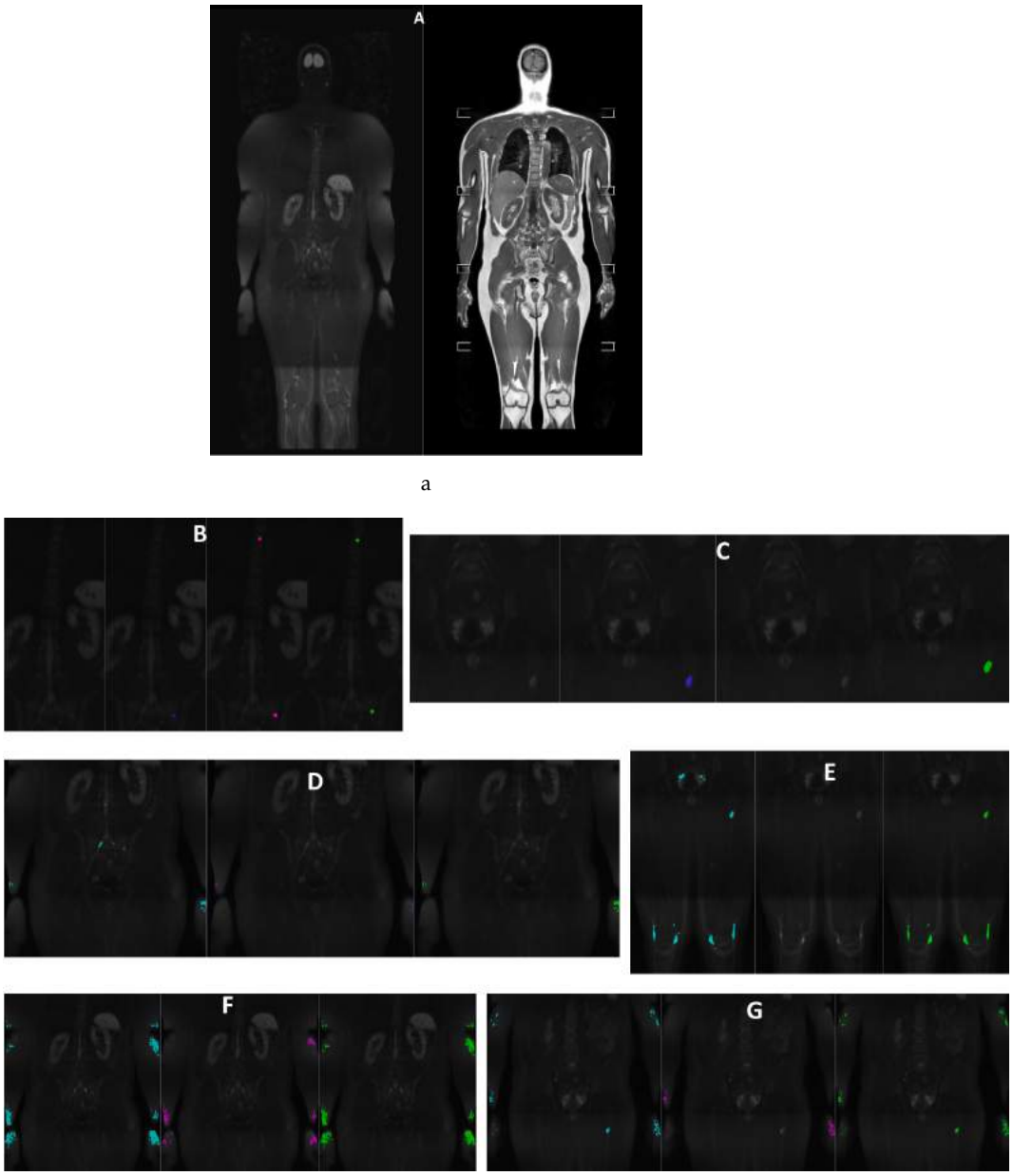
Figure 6.10: Side-by-side comparison of manual and smart segmentation of Case 1 (male). (a) Original DWI and T1w images: A shows a representative coronal slice of 800 s/mm^2 DWI and T1w images. (b) Segmentation: (B and C show two coronal slices of 4 experts' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 800 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by-voxel approach. Finally, the connected-component approach in light green. Visually, the connected component approach (D, E, F, G light green) shows the best agreement with the manual segmentation.

- **Scenario 2**

  Scenario 2 groups all WB images where at least one radiologist found no lesions (Cases 3, 4, 6, 7, 8, 10, 14, 17, 18, 19, 20). Explanation of why these images were not used to evaluate the overall mean of DSC is also provided.

  Case 14 and Case 3 are WB-MRI images acquired from the same patient, with six months apart. Since the results were similar for both images, only Case 14 segmentation results are presented (figure 6.11). Firstly, it is important to note that the normally hyperintense organs were successfully removed. It is also shown that one of the radiologists did not find any lesion, while three segmented a small lesion in the iliac bone, two segmented a lesion on the femur and on a vertebra. This shows that, generally, radiologists have low agreement on defining lesions, which is confirmed by the mean DSC between them (0.136). The results of the smart algorithm applied to the 1000 s/mm^2 b-value image show that neither the voxel-by-voxel or connected-component approach (using percentile 5th) detected the lesion on the iliac bone, although few pixels were detected on the Final Result DWI using k=3. This means that not even 5% of this small lesion's area is smaller than that of the muscle, failing one of the MM lesion criteria. The femur lesion, on the other hand, is detected by the connected-component but not by the voxel-by-voxel approach. However, this algorithm is also segmenting small areas of the arms and knees. As was already discussed, the arms are bright on DWI due to the use of surface coils. With the connected-component approach, most of them become one unique label which lie on muscle regions on T1w, thus have a typical muscle intensity. The knees, on the other hand, are bright on DWI because the station that contain them is wrongly hyperintense. This was an acquisition problem, often identified on several images, that can decline the overall algorithm performance (mean DSC 0.007). However, since the goal was not to build an autonomous algorithm, this can be easily identified and removed by the radiologist. The algorithm applied to the 1500 s/mm^2 b-value image shows lower performances (mean DSC 0.005), since it detects higher areas of the arms. Similarly, the femur lesion is not detected by the voxel-by-voxel approach.

  The difference between the DSC of the smart segmentation using the connected component approach on 1000 s/mm^2 image (0.007) and the manual segmentation (0.136) is justified based on the fact that one radiologist (E1) did not find any lesion and the algorithm delineated structures that are not lesions (arms, knees). Thus, this case will not be included to estimate the overall performance of the algorithm (mean DSC), since it is questionable whether the segmented area is a lesion or not. Plus, the fact that not even the fifth percentile can detect one of the lesions, may help to describe it as a non-lesion.



a

b

Figure 6.11: Side-by-side comparison of manual and smart segmentation of Case 14 (male). (a) Original DWI and T1w images: A shows a representative coronal slice of 1000 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 experts' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel--by-voxel approach. Finally, the connected-component approach in light green. The connected component approach applied to the 100 s/mm^2 image (D and E light green) shows, once again, the best agreement with the manual segmentation.

Besides Case 14 and Case 3, there were also tested other DWI images where at least one radiologist did not find any lesion: Case 4 (E1, E2); Case 6 (E1); Case 7 (E1, E2, E3); Case 8 (E1, E2); Case 10 (E2, E1); Case 17 (E1); Case 18 (E1, E4); Case 19 (E3); Case 20 (E1, E3, E4)[1].

For instance, although three radiologist found possible lesions on Case 6, none of them is concurrent: the segmentation were all different. This lack of agreement is very problematic and this tool arises from the need to solve this. In this case in particular, the incorporation of T1w information, either using voxel-by-voxel or connected-component with a percentile of 1, did not segment any lesion on the skeleton. This is in accordance with the radiologist that did not find any lesion and with the disagreement found between the other three. Still, part of the arms was again detected. However, since there are three radiologists that detected lesions, even all different, the DSC of table A.1 is referent to using only the Final Result DWI with k=3, since this comprised some of the segmented areas.

Similarly to Case 6, two radiologists did not found lesions on Case 8. However, the ones that found possible lesions have very low agreement: one segmented a region in the femoral head and a small area in the iliac bone, while the other only segmented the same small area in the iliac bone (figure 6.12-B). The algorithm fails to detect these lesions, either using the 1000 or 1500 s/mm^2 b-value, considering k= 1.5 or 3.

This small iliac lesion is a typical lesion when bone marrow biopsy is performed on MM patients to assess the percentage of bone marrow infiltration. This spot, hyperintense on DWI and dark on T1w (figure 6.12-A), can be retained for months after the procedure. In this patient in particular, the biopsy was done 2 weeks before acquiring the MRI images. Thus, it is not a real MM lesion, although segmented by two radiologists. This adds value to this smart algorithm since it seems to be able to distinguish between lesions left by a needle from real MM bone lesions.

Case 20 is part of a series of three WB-MRI images from the same patient, over time. The first one is Case 19 before treatment. After 10 months and during treatment, the patient did the Case 20. After 12 months, Case 21 was performed. E3 did not find any lesion on Case 19, while the others found small ones. Then, E3 found a small lesion on Case 20 that was not detected by the other three radiologists (they detected other small areas). Finally, on Case 21 only E2 found lesions. Although all the radiologists found possible lesions on Case 20 and they were correctly detected by the Final Result DWI on 1000 and 1500 s/mm^2 b-values, using k=3, no skeleton lesion was detected when incorporating T1w information. The voxel-by-voxel and connected-component approaches were experimented, and not even using a percentile of 1, with the last approach, was possible to find lesions. This means

[1] In brackets are the radiologists that did not found lesions.

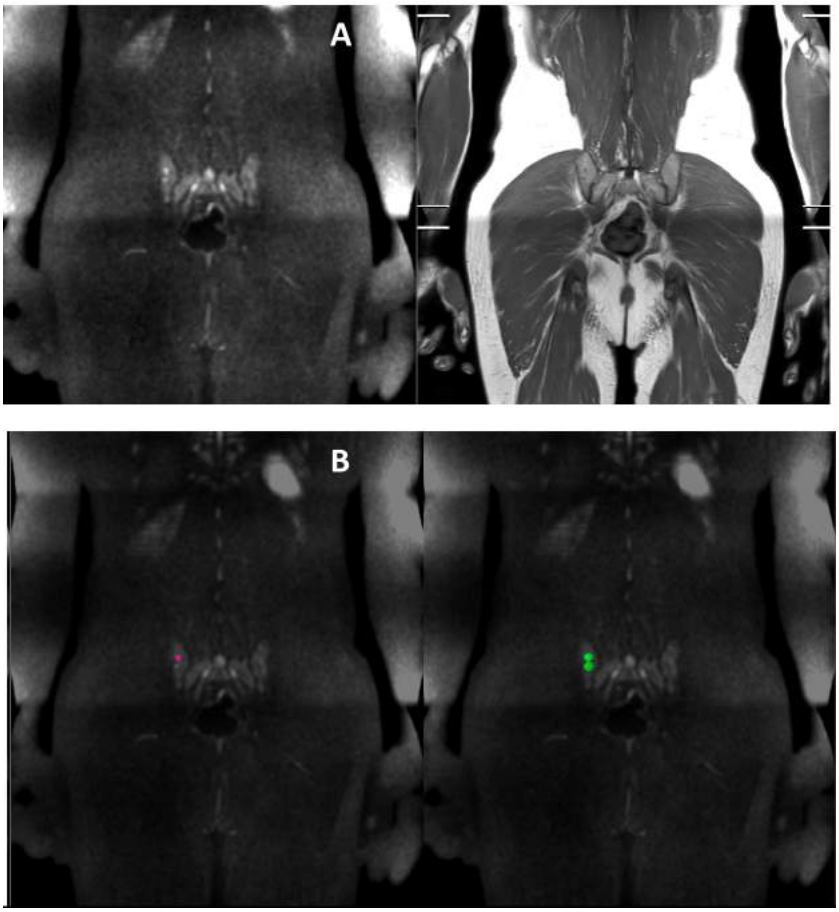


Figure 6.12: Possible lesion on the iliac bone segmented by E3 (red) and E4 (green) on Case 8 (male). A- representative coronal slice is zoomed to fit the iliac bone on DWI (left) and T1w (right). There are spotted two bright points on DWI that are coherent with dark points on T1w. B- manual segmentation of E3 and E4.

that not even 1% of the area of each connected lesion was darker than the muscle, failing one of the criteria discussed in section 4.2.1. This could be associated to the good clinical outcome after treatment, since the patient achieved a very good partial response to treatment.

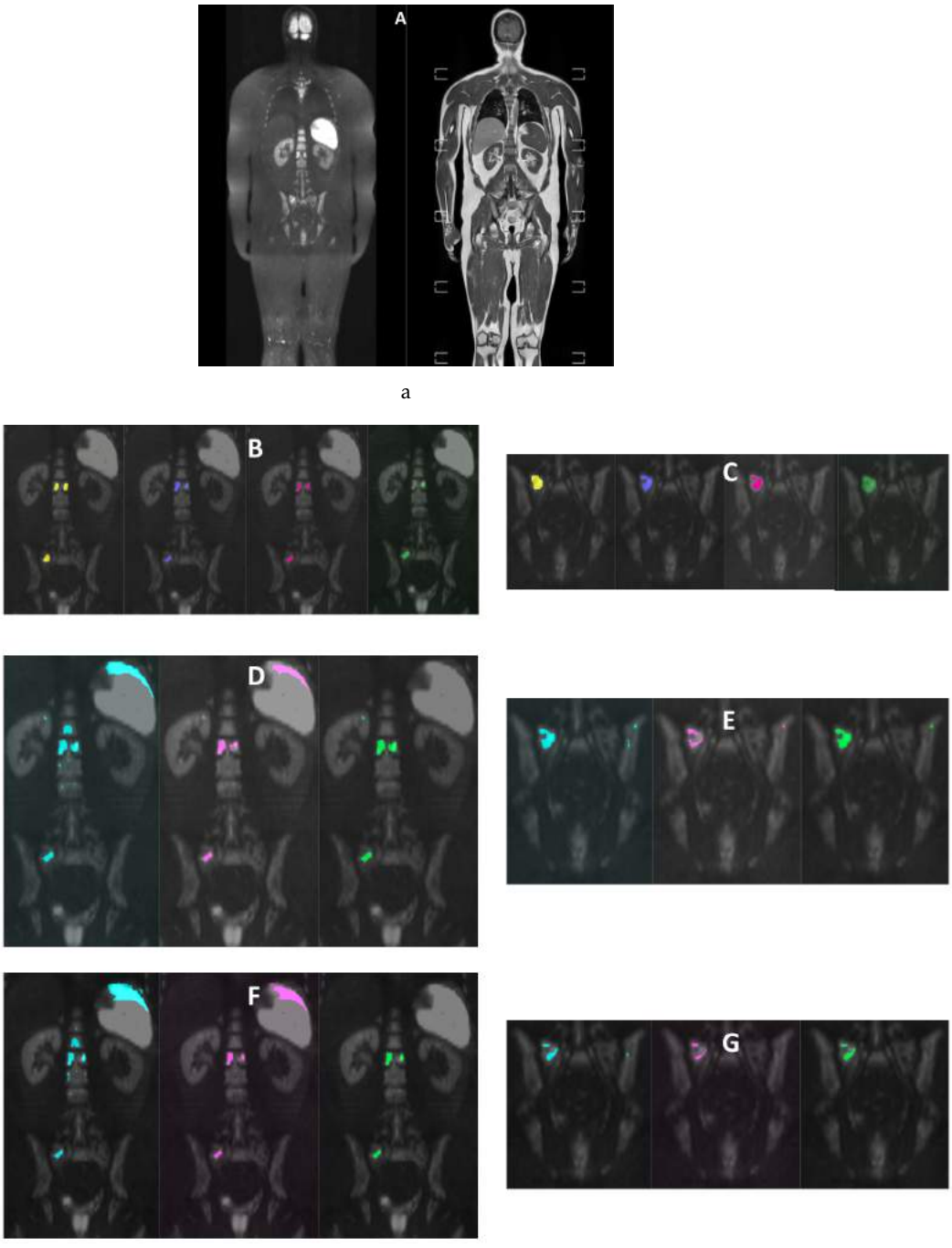
- **Scenario 3**

  In this scenario, a diffuse pattern is presented (Case11). Radiologists segmented the more hyperintense regions of the skeleton. In this case, the spleen was not completely removed by the algorithm because it appears to be larger than normal. By complementing the connected-component approach with the T1w image, this problem was solved.

  Figure 6.13 shows several comparisons between manual (B and C) and smart segmentation (D, E, F and G) of Case 11. The four radiologists agree on the number of lesions (B and C), although the areas are not exactly the same (mean DSC 0.448). Although not displayed on figure 6.13, there was also reported by all radiologists a small lesion on the iliac bone. One of them (E3) found more than one lesion there, which explains the smallest DSC coefficient between E3 and the other radiologists, as can be verified on appendix A.

  The Final Result DWI using the 800 s/mm^2 b-value image (D and E) applied k=4, while k=3 was used to obtain the Final Result DWI for the 1500 s/mm^2 image. Once again, these k values were the optimal ones that covered all lesions segmented by the radiologists. Both Final Result DWI using the 800 and 15000 s/mm^2 b-value image did not successfully remove the spleen. This might suggest that the registration algorithm should be optimized for hyperintense organs that may have large sizes, as the spleen, which is in agreement with the findings of section 6.1. As expected, the voxel-by-voxel approach was not able to remove that area of the spleen, since it appears darker than the muscle in T1w. However, it detects the vertebra and sacro's lesions on both b-values, meeting the radiologists' segmentation. The connected--component approach correctly removes the spleen using a percentile of 55 and detects the vertebra and sacro's lesions. Also, no hyperintensities were segmented on the arms. This can be explained based on the fact that this patient's lesions are very hyperintense on DWI (figure 6.13 A- left), outstanding from the smooth hyperintense areas in the arms.

  The small iliac bone lesion segmented by the radiologists was also detected by the Final Result DWI but not detected when incorporating T1w information (voxel--by-voxel and connected-component), using the percentile of 55. However, it was detected by lower percentiles (below 15th), together with other areas, which shows that small iliac area has a T1w intensity very different from that of the other detected lesions. This suggests that it may not be a lesion. This small disagreement justifies the fact that the mean DSC of the manual segmentation using the 800 s/mm^2 b--value image is just over half than that of the manual (0.254) and that the mean PPV is 0.178.



a

b

Figure 6.13: Side-by-side comparison of manual and smart segmentation of Case 11 (male). (a) Original DWI and T1w images: A shows a representative coronal slice of 800 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 800 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel--by-voxel approach. Finally, the connected-component approach in light green. Visually, the connected-component approach applied to the 800 and 1500 s/mm^2 image are very similar, although E (800 s/mm^2) appears to cover a larger area, showing better agreement with the manual segmentation.

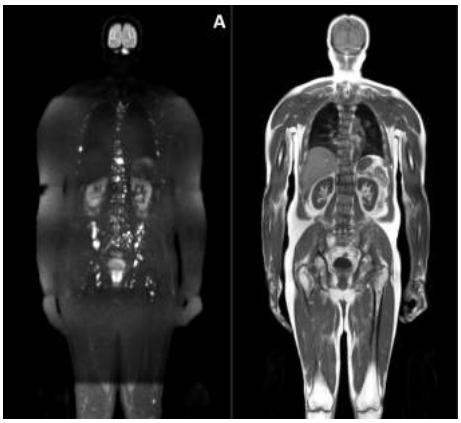
- **Scenario 4**

  This scenario represents an advanced case of a focal pattern, which could be misinterpreted as diffuse (Case 15). Lesions are present all over the skeleton, which makes a very hard task to manually assess tumor burden.

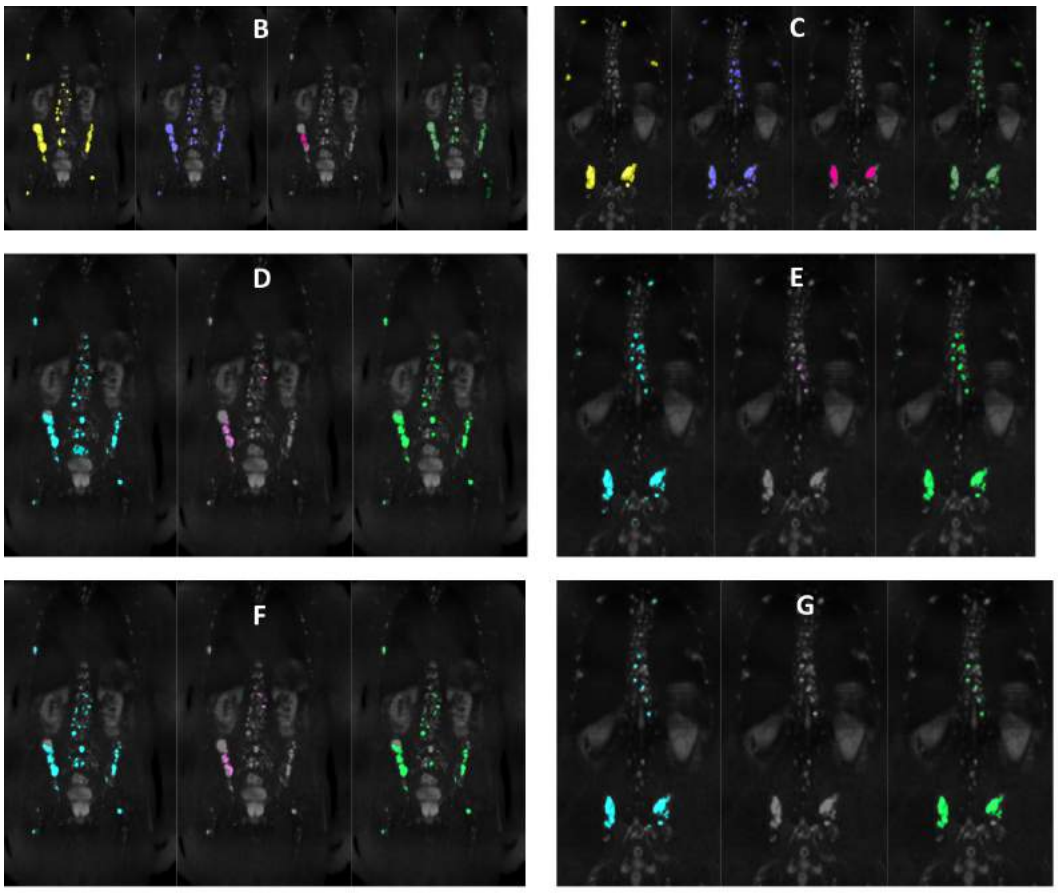
  Figure 6.14 shows a perfect example of a patient with many possible lesions, easily distinguishable in the DWI image (figure 6.14-A left). Visually, the manual segmentation seem to be very similar, except for the red (E3) one. A more careful look reveals that there are some areas that were not segmented by all radiologists, especially the vertebras on the slice shown in C. This justifies the fact that the mean DSC of the radiologists' manual segmentation is 0.412.

  A k value of 3 was enough to cover all lesions both on 1000 and 1500 s/mm^2 b-values DWI images (blue D and F), while the connected-component needed a percentile of 1 to guarantee that all lesions would be included. We decided to use this percentile since half of the radiologists'segmentation are very similar to that of Final Result DWI, so a small percentile would include most of the lesions found on diffusion. The voxel-by-voxel approach fails to detect most of the lesions, since it searches from the voxels that are darker than the muscle, which is impossible to be done visually. The radiologist visually decides whether the intensity of a specific area is similar to that of a muscle; if it is, they consider it a lesion; if it is not, they can discard it. The connected-component approach using the 1000 s/mm^2 DWI image showed very good agreements with the manual segmentation one, resulting in a mean DSC of 0.372, mean PPV of 0.340, mean NPV of 1.000 and sensitivity of 0.580.

  An advantage of this smart segmentation algorithm is the time of processing. In this case in particular it is a plus, since the radiologist took, on average, one hour to study and segment these images, while the algorithm took 10 minutes. This does not mean that the smart algorithm is better than the radiologist, it means that this can be used as a tool to decrease the overall examination times.



a



b

Figure 6.14: Side-by-side comparison of manual and smart segmentation of Case 15 (male). (a) Original DWI and T1w images: A shows a representative coronal slice of 1000 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel--by-voxel approach. Finally, the connected-component approach in light green. Visually, the connected-component approach applied to the 1000 and 1500 s/mm^2 image are very similar, although E appears to cover some more lesions than G, which were also detected by 3 radiologists.

- **Scenario 5**

  This scenario groups two possible cases of bone marrow reconversion (Case 16 and Case 12), both with diffuse patterns, and two cases with very few focal lesions (Case 5 and Case 9).

  Figure 6.16 (Case 16) shows a particular example of what might be bone marrow reconversion (from yellow to red). The DWI image (figure 6.16-A left) shows some hyperintense spots on the femur, vertebras and iliac bones that could be interpreted as MM bone lesions. On the other hand, T1w image shows that the axial and appendicular skeleton have dark intensities, very similar to that of muscle. It is also visible that the hyperintense lesion seen on the DWI image, is showing darker intensities in the T1w than the surrounding bone area.

  Considering B and C, we can see that the radiologists' opinion is diverged: while E3 (red) considers almost the whole skeleton as lesion, the other three only segmented small areas. Again, this problem raises many questions: why do radiologists, with almost the same level of expertise, classifying lesions in such different ways? In this case, the mean DSC of their segmentation is 0.101.

  The smart algorithm applied to the 1000 s/mm^2 DWI image and using k=3.75 and a percentile of 35 (D and E) shows high agreement with the majority of the manual segmentation, when compared with the 1500 s/mm^2 DWI image (F and G), using k=3 and percentile of 35. However, the mean DSC (0.040) does not confirm this, since the algorithm also segmented part of the legs (figure 6.15). As seen in A, the last station that covers the legs in DWI is wrongly hyperintense. This contaminates the image and the algorithm is mistaken when classifying it as outlier. This problem was also present on Case 13, which justifies the small mean DSC of the smart algorithm (0.008) versus the manual one (0.155). Nevertheless, the radiologist could easily remove those areas after running the algorithm, obtaining a fast and accurate segmentation.

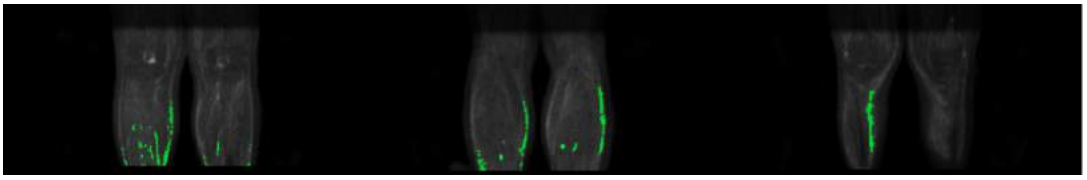


Figure 6.15: Several coronal representative slices of the legs of Case 16 (male), using the connected-component segmentation applied to the 1000 s/mm^2 image. The segmented areas are hyperintense on DWI and 35% darker than the muscle on T1w. However, it is easily seen that these areas are not near the skeleton, thus would never be mistaken by MM lesions.

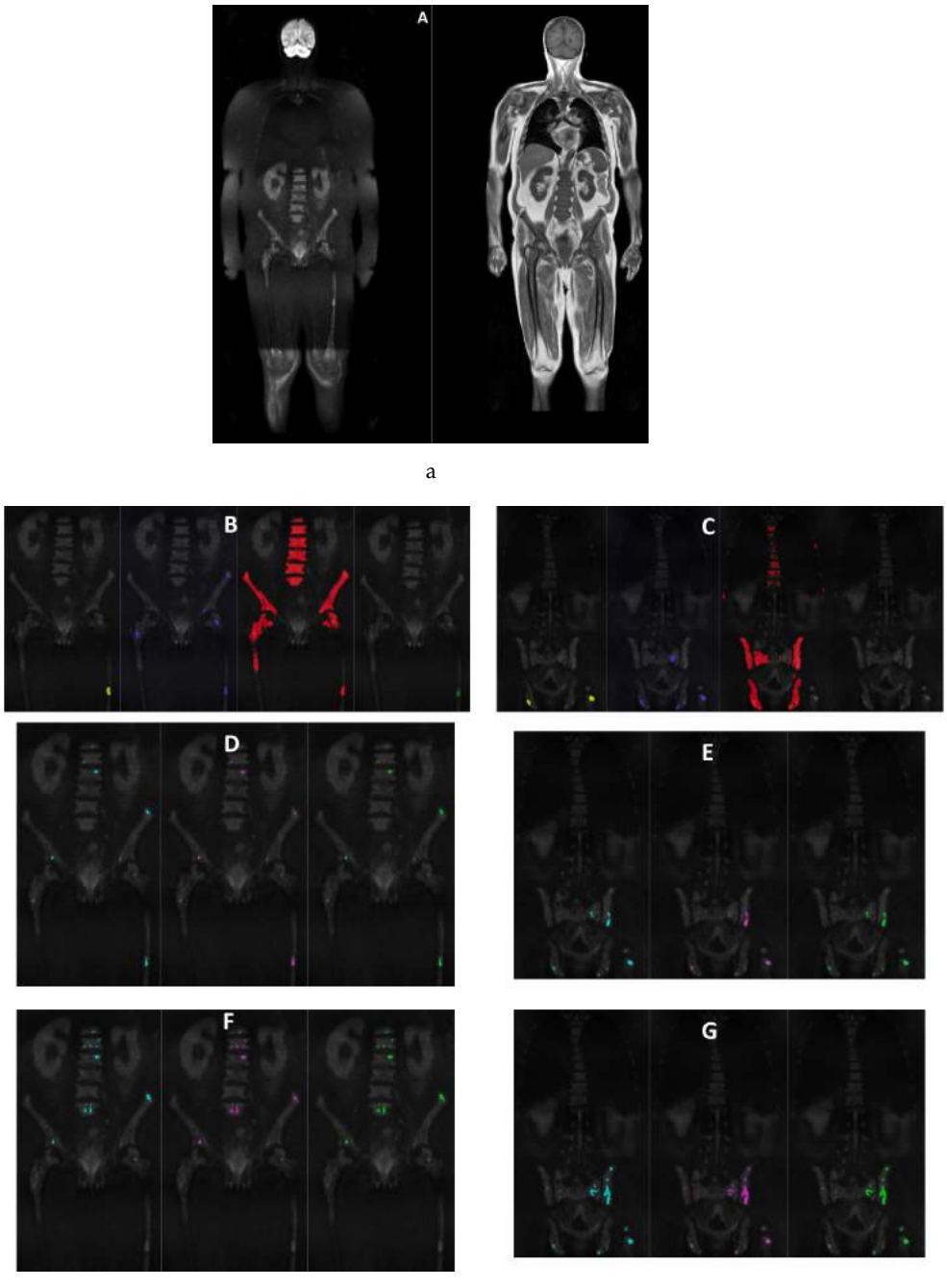


Figure 6.16: Side-by-side comparison of manual and smart segmentation of Case 16 (male). (a) Original DWI and T1w images: A shows a representative coronal slice of 1000 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by--voxel approach. Finally, the connected-component approach in light green. The manual segmentation are very different, revealing high disagreement between the radiologists. Visually, the connected-component approach applied to the 1000 s/mm^2 image shows better results than the 1500 s/mm^2 image, since this last one is covering small areas that were not detected by the three radiologists that show better agreement (E1, E2, E4).

The same problem happened on Case 12: one of the radiologists segmented almost the whole skeleton (E1), while the other only detected small lesions that were not consistent between them. We verified that almost the whole skeleton would be segmented by using k under 4 so, in order to agree with the majority of the radiologists, k=5 was used. Once again, the voxel-by-voxel approach failed to detect the lesions both on 800 and 1500 s/mm^2 DWI images. The connected-component approach used a percentile of 8 to cover all the small lesions, although some other areas were also segmented (for instance, vertebras that were segmented by E1). Still, the connected-component approach applied to the 800 s/mm^2 showed higher agreements (mean DSC 0.077) than the manual one (0.036).

It is verified that the smaller the number of lesions found by the radiologists, the smaller the agreement between them. Case 5 (figure 6.17) and Case 9 are examples of this. Figure 6.17 shows some zoomed coronal slices of Case 5 that show the small agreement between radiologists: on slice A only two radiologists segmented small lesions (E2, E1) and on B and C again only two detected lesions (E2, E4), resulting on a mean DSC of 0.057. The smart segmentation algorithm covered all this lesions on Final Result DWI using k=3 both on 1000 and 1500 s/mm^2 DWI images. However, no lesion was detected on the skeleton when applying the T1w approaches, not even using a very small percentile.

DSC, PPV, NPV and sensitivity were computed for all approaches (voxel-by-voxel and connected-component), using both the highest and the computed b-value DWI images. As described previously, only the statistics of those where all radiologists found lesions will be used to assess the overall performance of the algorithm. The others are displayed in appendix A.

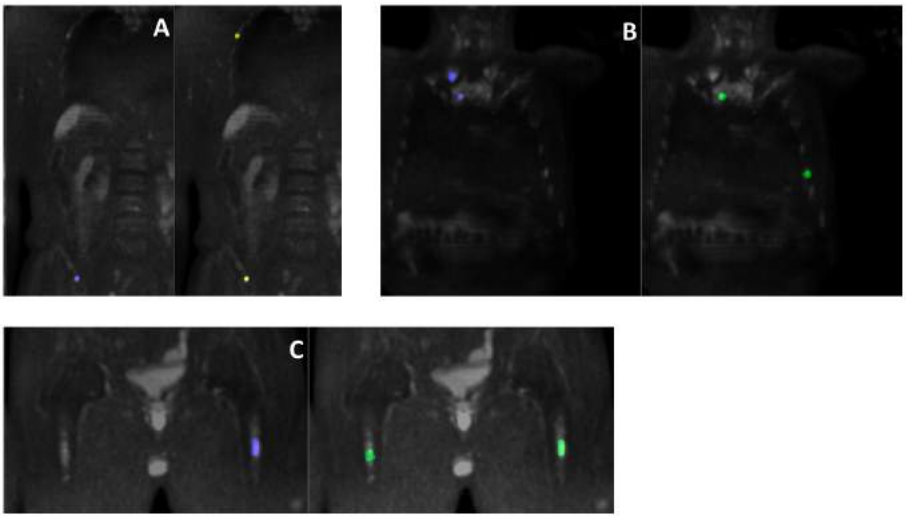


Figure 6.17: Representative coronal slices of the manual segmentation of Case 5 (male). On slice A, only two radiologists found lesions (E2, E1). On B and C only E2 and E4 found lesions. E3 found one lesion on the sternum.

Female

- **Scenario 1**

  This scenario presents and discusses Case 1 and Case 12. The first has a variegated pattern. The second is an example of a diffuse pattern.

  Figure 6.18 shows some representative coronal slices of Case 1, where all radiologists agree on the shoulder's lesion (B) but show some disagreement on some areas of the left femur (C). The mean DSC of their segmentation is 0.351.

  It is also noticeable that the atlas algorithm is correctly removing the normal hyperintense organs, since these structures were not segmented by the Final Output DWI (either using the 1000 or 1500 s/mm^2 b-value image).

  As expected, by using only the Final Output DWI algorithm (D, E, F, G) the result shows over segmentation of the arms and legs, especially using the 1500 s/mm^2 image. As explained for the males' scenarios, this is due to the use of surface coils when acquiring the images, whose intensity becomes more evident by increasing the b-value. Plus, the last station that covers the legs is, again, wrongly hyperintense. Although both 1000 (k=3) and 1500 (k=1.5) s/mm^2 identify the main lesion detected by the radiologists, the 1000 s/mm^2 shows better results, since has less influence from the arms and legs hyperintensities. k were chosen as those that detected the most similar areas to those from the manual segmentation. When incorporating T1w information, both voxel-by-voxel approaches (middle image of D, F, E and G) detect the shoulder's lesion but do not identify the femur one. This means that the shoulders area's voxels have an intensity lower than that of the muscle, contrary to the femur suspicious area. The connected-component approach using the highest b-value image (right image of D, E) and 55% percentile shows high agreement to the radiologists, segmenting the shoulders' area and a small area of the left arm. Consequently, the mean DSC is 0.303, very close to that of the manual segmentation (0.351). Plus, the extracted statistic metrics show greater similarity to E1 and E2, confirmed by the PPV: 0.439 and 0.467, respectively.

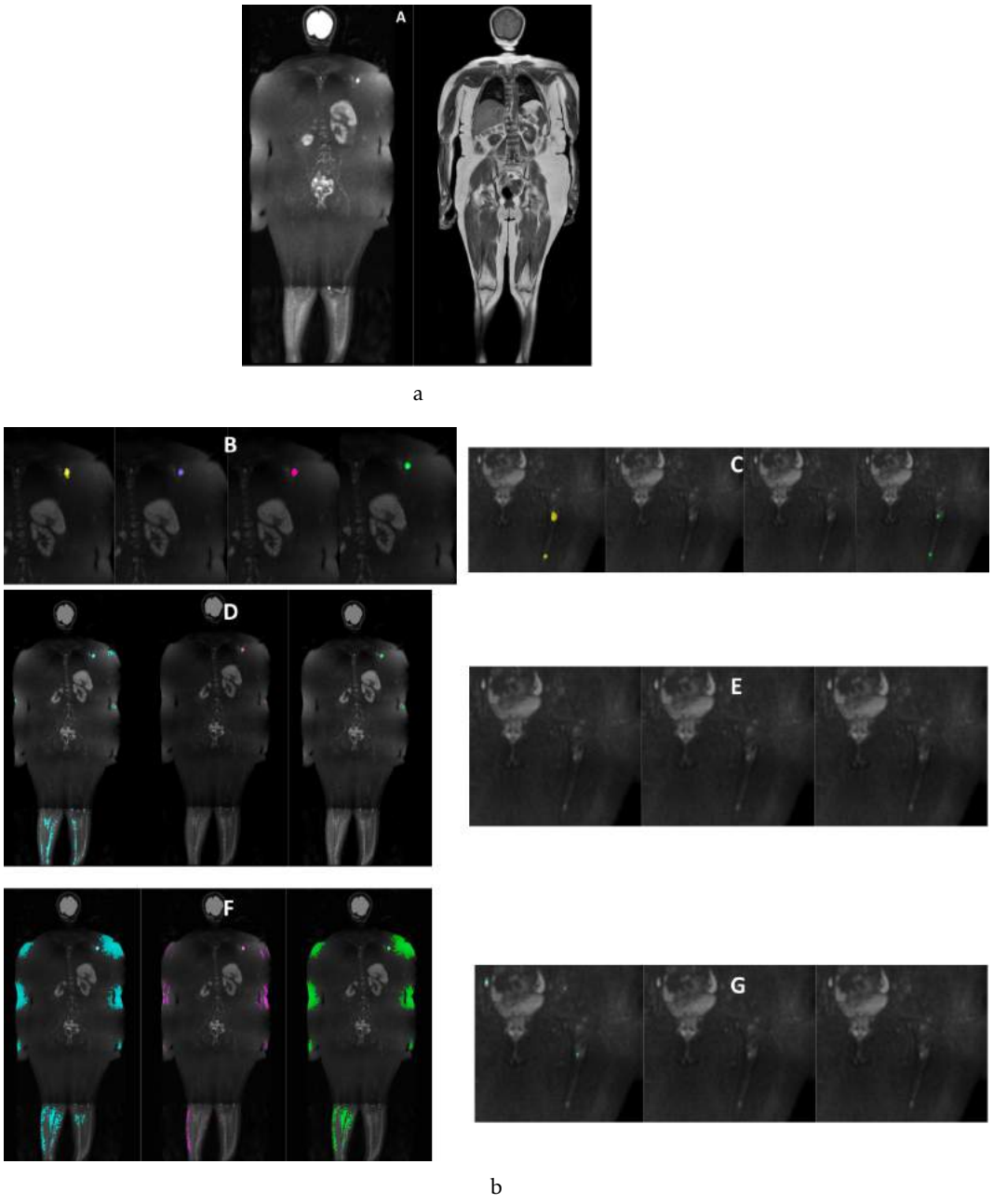


Figure 6.18: Side-by-side comparison of manual and smart segmentation of Case 1 (female). (a) Original DWI and T1w images: A shows a representative coronal slice of 1000 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by--voxel approach and the connected-component approach in light green. All radiologists detect the same lesion in the shoulder's area, but two disagree on segmenting an area in the femur. The connected-component approach applied to the 1000 s/mm^2 image shows better results than the computed one, since this last one is covering the arms and legs. The voxel-by-voxel approach also shows good agreement to the manual segmentation, reveling that the suspicious shoulder' area is, most probably, a lesion.

Figure 6.19 shows another example of a painstakingly slow segmentation (Case 12), since there are reported several suspicious areas which require a close eye. Once again, a high degree of disagreement is found between radiologists' segmentation, resulting in a mean DSC of 0.383. The main deviations are the iliac bone and vertebras, which were mainly segmented by E1 (yellow). This disagreement can be explained due to the nature of the pattern (diffuse).

Once again, it is shown in sub-images D, E, F and G that the registration algorithm is suitable for different body shapes, enabling the correct removal of the normal hyperintense organs. Without this, the outlier segmentation would include them.

As seen on sub-image A, the DWI is smoothly affected by the surface coils. This effect is decreased when applying the Final Output DWI because this case has a lot of hyperintense suspicious areas, being accentuated over the arms. Thus, the smart algorithm's result includes only small areas of arm hyperintensities, more visible using the computed b-value image (F and G). k value of 3 was enough to detect the suspicious areas on both b-values. Both voxel-by-voxel approaches are failing some regions that were segmented by all the radiologists. The connected-component approach applied to the 800 s/mm^2 b-value image is showing higher agreements (mean DSC 0.451, mean PPV 0.448, mean NPV 1.000, mean sensitivity 0.593) than the radiologists, which stresses the value of this smart algorithm. Plus, it is to be noted that this tool takes 10 minutes to get a segmentation, including DWI and T1w information.

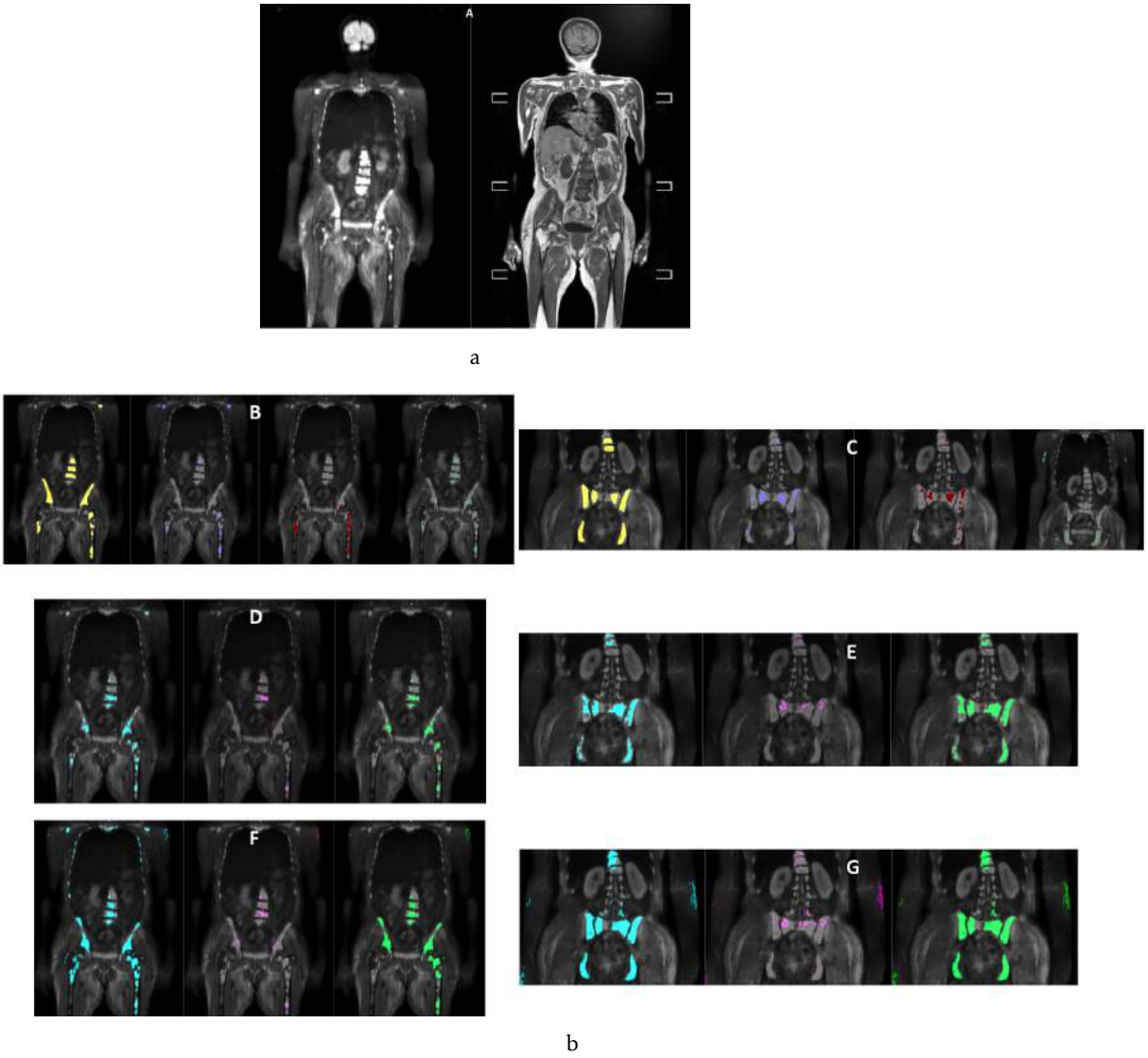


Figure 6.19: Side-by-side comparison of manual and smart segmentation of Case 12 (female). (a) Original DWI and T1w images: A shows a representative coronal slice of 800 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by-voxel approach. Finally, the connected-component approach in light green. The connected-component approach applied to the 800 s/mm^2 image shows high agreement with the manual segmentation.

- **Scenario 2**

  This scenario presents and discusses Cases 13, 14 and 15, which are a series of MRI images from the same patient over time. The first one is before treatment and presents a diffuse pattern, which evolved to a focal pattern after treatments.

  Figure 6.20 shows Case 13, very similar to Case 12, where a big area of lesions is segmented by all the radiologists. However, the areas are not all coincident. As displayed in sub-images B and C, there are radiologists who might be over or sub segmenting. This divergence results in a mean DSC of 0.331. Their disagreement is probably due to uncertainties on differentiating between disease and marrow reconversion. However, since this case is before a first line of treatment, all findings are most likely to be MM lesions. Regarding the smart segmentation algorithm, it is clear that the normal hyperintense organs are correctly removed: although not always visible on the chosen slices, the segmentation is not including the brain, kidneys, spleen and bladder (sub-images D, E, F and G). As expected, for the reasons already discussed, the connected-component approach applied to the 800 s/mm^2 b-value image (D and E) is the one that achieves best results. In fact, its agreement is greater than that of the radiologists, reaching a mean DSC of 0.413. To meet the radiologists' segmentation, a percentile of 1% was used.

  Figure 6.21 shows Case 14 performed during treatment. Manual segmentation show good agreement (mean DSC 0.537), indicating suspicious areas on the iliac, femur, ribs and vertebras. Regarding the smart segmentation, it is very clear that by using the 1500 s/mm^2 the results are much worse. The influence of the arm hyperintensities is much accentuated, even though this segmentation also includes the reported suspicious areas (mean DSC 0.035). On the other hand, this segmentation improves the removal of the normal enhancing organs, as the kidneys, whose intensity is decreased with high computed b-values (section 6.2.1). Observing the results of using the 800 s/mm^2 b-value image, some regions of the kidneys and bowels were not eliminated (sub-image E), even including the voxel-by-voxel and connected component approaches. Nonetheless, the connected-component approach successfully segments the areas of the vertebras, ribs, and iliac bone by using k= 2.5 and 22% percentile. Although showing a relative inferior performance (mean DSC 0.364, mean PPV 0.389, mean NPV 1.000, mean sensitivity 0.361) to that of the manual segmentation, it has potential to be improved by removing the areas that are obviously not MM lesion (e.g. kidneys and bowels).

  The last image of these series of exams is Case 15, where very small areas were reported by the radiologists. Its mean DSC was 0.420. This patient achieved very good partial response to treatment, also confirmed by the decreased number of lesions between pre (Case 13) and post therapy (Case 15). The smart segmentation

algorithm using the outlier segmentation in DWI with k=2.5 detects some of the reported suspicious areas but not all. If the k value is decreased, the segmentation will include hyperintense areas that are not lesions (bowels). Thus, k=2.5 was a commitment between finding the greater number of lesions and discarding other hyperintense areas. When incorporating the T1w information, the DWI detected lesions remained. However, this case shows a very low agreement between the manual segmentation (mean DSC 0.099).

- **Scenario 3**

  Case 16 and 17 are WB-MRI images with an interval of six months, from a patient that progressed from MGUS to MM. Both cases are considered variegated pattern, with the progressed one showing a very inhomogeneous bone marrow with numerous small lesions.

  As expected, the biochemical evolution is also confirmed by these MRI exams: Case 17 (figure 6.22) shows a higher tumor burden than Case 16. For reasons of content limitation, only Case 17 will be shown.

  Figure 6.22 shows, once again, that the algorithm is suitable for removal of the normal hyperintense organs, fulfilling one of the goals. Once again, the manual segmentation of the radiologists lacks agreement (mean DSC 0.379). Regarding the smart segmentation, although not optimized, is covering most of the suspicious areas using the 1000 and 1500 s/mm^2 b-value images. However, the 1500 s/mm^2 is influenced by the hyperintensities of the arms, contaminating the results. On the other hand, as shown in A, the 1000 s/mm^2 also contains a wrongly hyperintense station. This results on detecting the legs when using this image, since it is considered as an outlier on DWI. Part of it was removed by incorporating T1, but the major part was not. This can explain why the mean DSC of the connected-component approach of the 1000 s/mm^2 is less than half of the manual one (mean DSC 0.112) and why the PPV is low.

  Cases 2 and 9 were very similar to Case 17. Their mean DSC are both less than that of the manual segmentation but show a low false negative rate with at least one of the radiologists. For instance, in Case 2, radiologists have low agreement between them (mean DSC 0.200). The smart segmentation algorithm has a mean DSC of 0.116 (more than half of the manual segmentation one) but shows higher similarity to E1 (DSC 0.170, sensitivity 0.768). In fact, this smart algorithm could help E2 to have its segmentation closer to E1, since its mean DSC is less than 0.116 (0.089) and its sensitivity is 0.219.

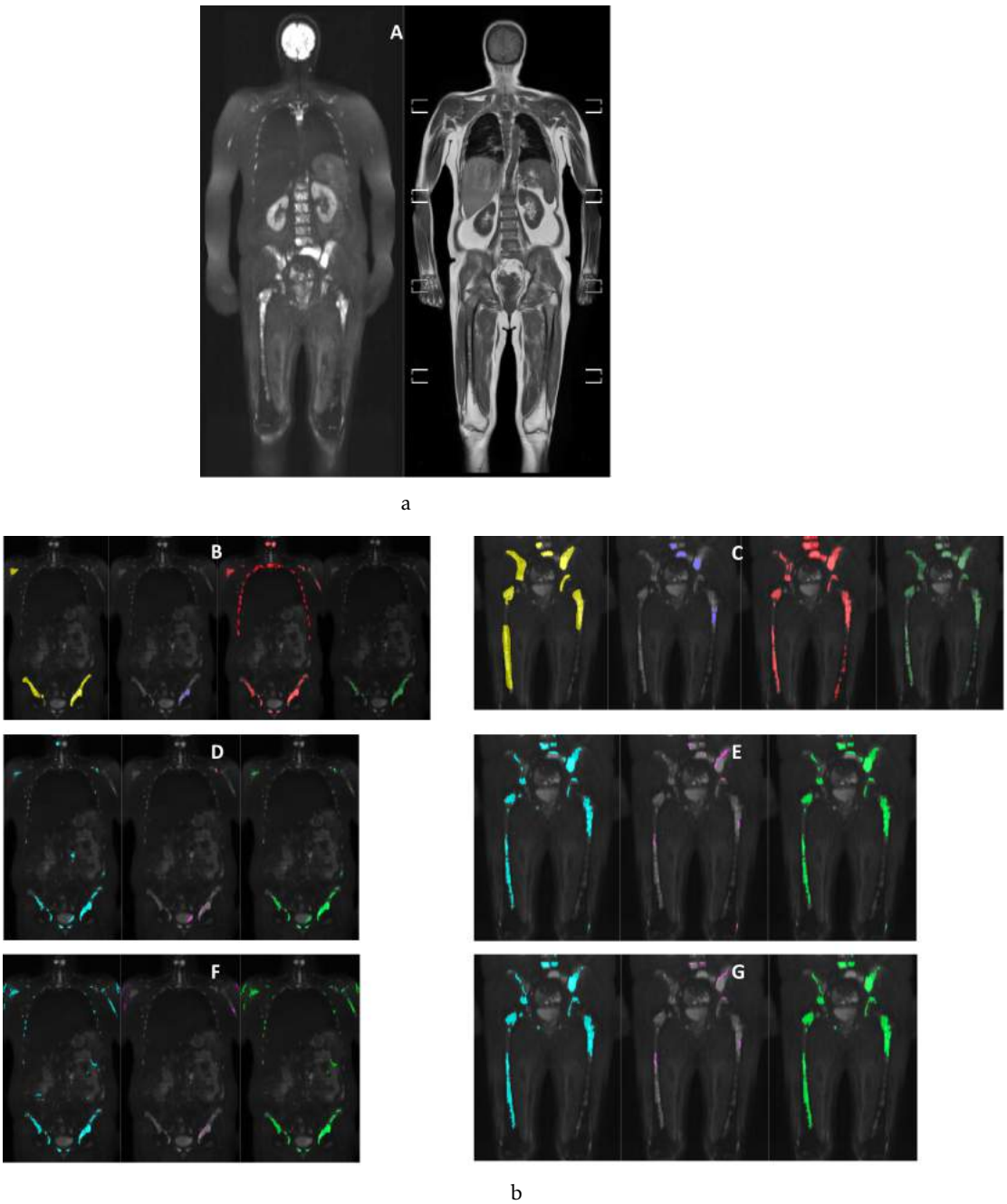


Figure 6.20: Side-by-side comparison of manual and smart segmentation of Case 13 (female). (a) Original DWI and T1w images: A shows a representative coronal slice of 800 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by-voxel approach. Finally, the connected-component approach in light green. The connected-component approach applied to the 800 s/mm^2 image shows greater similarity than manual segmentation.

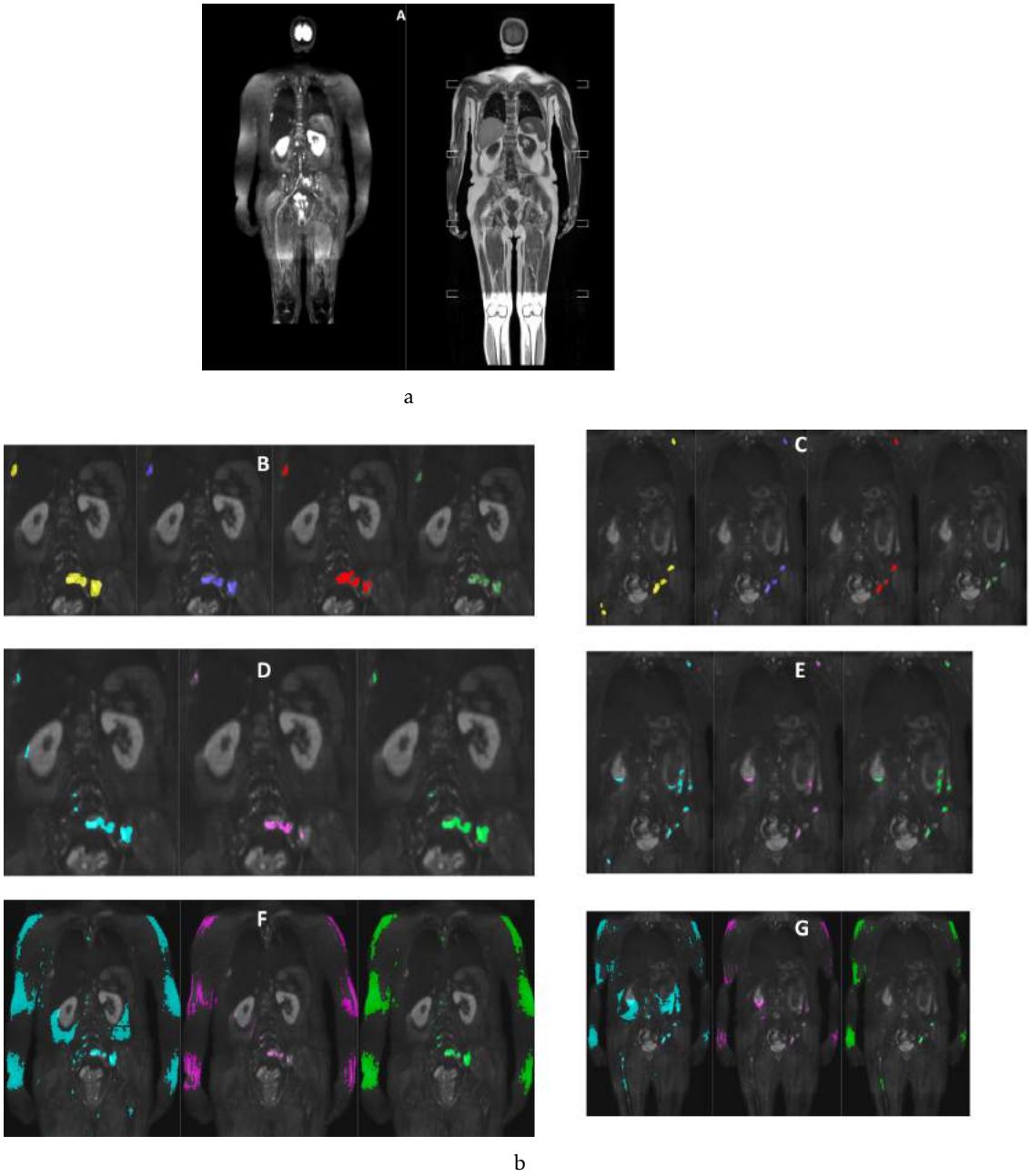


Figure 6.21: Side-by-side comparison of manual and smart segmentation of Case 1 (female). (a) Original DWI and T1w images: A shows a representative coronal slice of 800 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 800 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by-voxel approach. Finally, the connected-component approach in light green. The connected--component approach applied to the 1000 s/mm^2 image is detecting most of the lesions segmented by the radiologists but is also detecting part of the right kidney and bowels. This worsens when applying the voxel-by-voxel or connected-component approach to the 1500 s/mm^2, which highlights the hyperintensities of the arms.

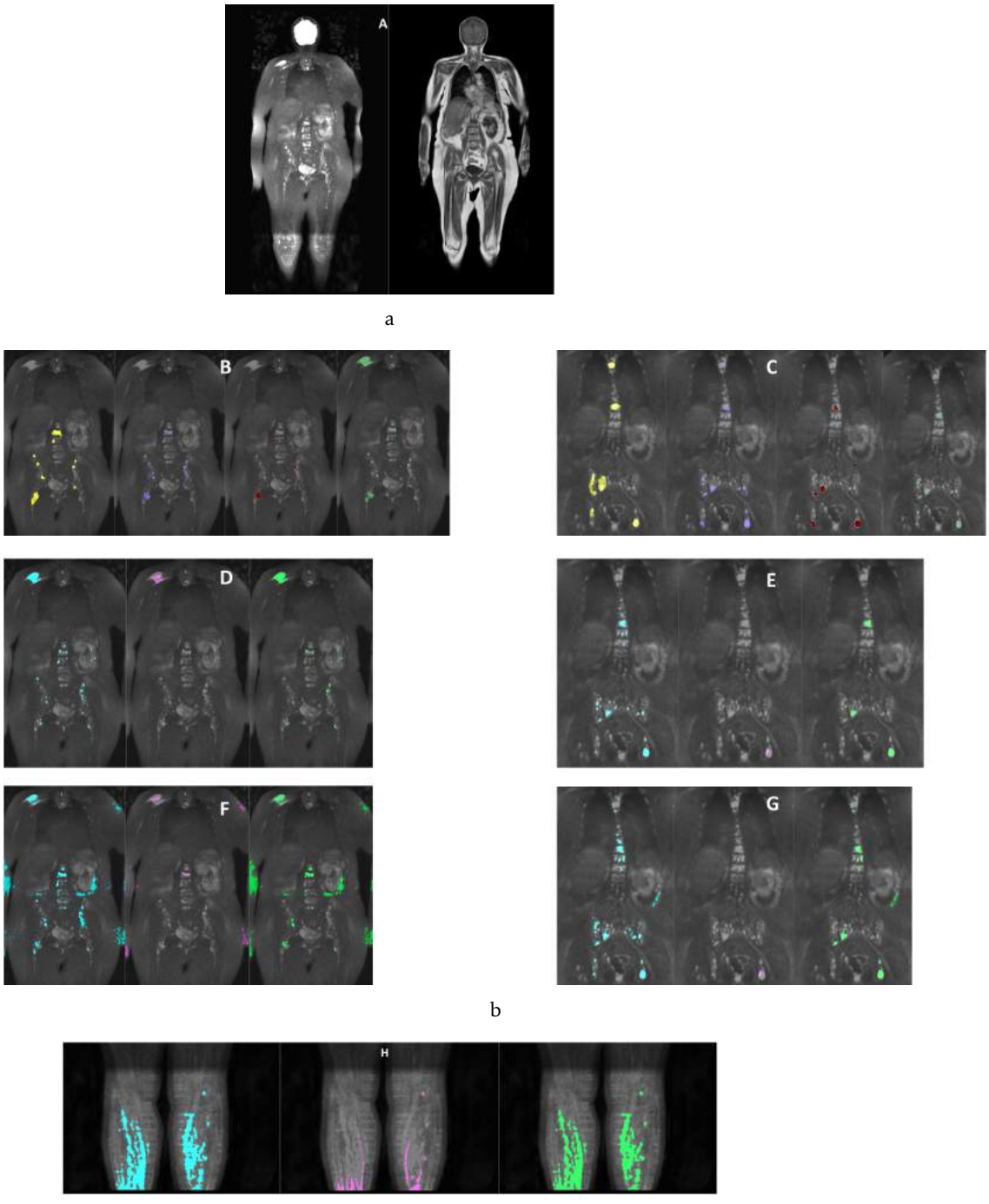


Figure 6.22: Side-by-side comparison of manual and smart segmentation of Case 17 (female). (a) Original DWI and T1w images: A shows a representative coronal slice of 1000 s/mm^2 DWI and T1w images. (b) Segmentation: B and C show two coronal slices of 4 radiologists' manual segmentation (yellow, purple, red and dark green). D and E show the same coronal slices using the smart segmentation tool applied to the 1000 s/mm^2 image. Similarly, F and G display the algorithm applied to the computed 1500 s/mm^2 DWI image. First, the Final Result DWI is displayed in blue, followed by the pink voxel-by-voxel approach. Finally, the connected-component approach in light green. The connected-component approach applied to the 1000 s/mm^2 image is detecting most of the lesions segmented by the radiologists but is also detecting a big area of the last station of DWI (H), that contains the legs.

- **Scenario 4**

  Similar to the male's Scenario 2, where at least one of the radiologists did not find lesions on DWI images, this also happened on Cases 3, 4, 5, 6, 7, 8, 10 and 11. There were no lesions detected on Cases 5, 6 and 10.

  On Case 3, only E2 found a small suspicious area on the ribs, that was not detected by the outlier segmentation (using k=3). Also, this small area's intensity was not smaller than that of the muscle on T1w image. The outlier segmentation on the 1000 s/mm^2 b-value DWI image, as expected, completely removed the normal hyperintense organs and did not segment any areas on the skeleton. However, as displayed on figure 6.23, sub-image A and B, it detected some small areas of spinal nerves, sinus liquid, nerves and part of the legs and knees (since the last station of DWI was wrongly hyperintense). Using the connected-component approach only part of the legs and sinus liquid remained (figure 6.23, sub-image D and E). Similar results were obtained by using the computed b-value, with the addition of the arms. As was discussed for the male's cases, if there is question if a DWI image has or not lesions, then this smart segmentation algorithm should not be used. This relays on the assumption that lesions are always stood out from the rest of the image. If there are no lesions, then other structures would be wrongly highlighted, for instance, sinus liquid and neck glans.

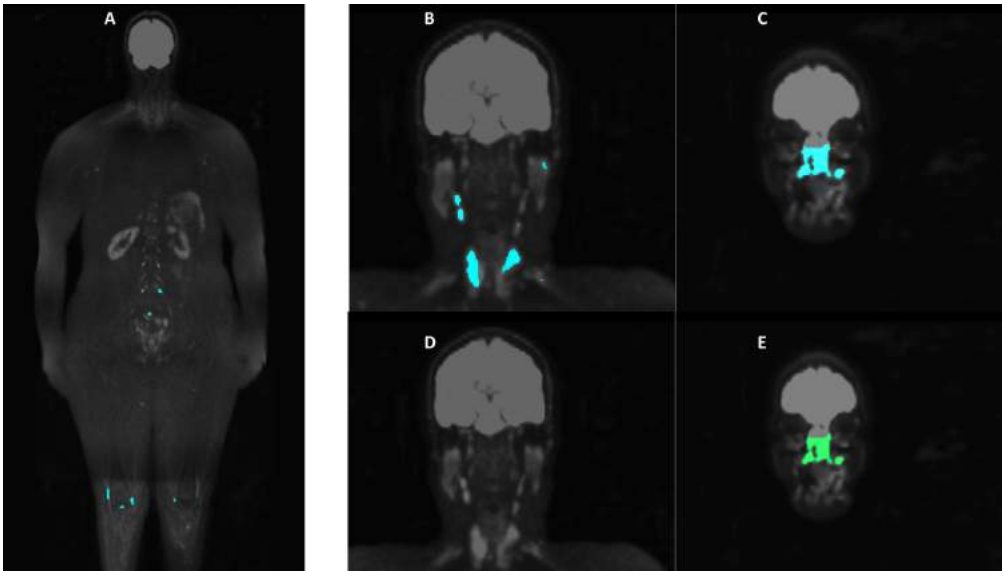


Figure 6.23: Smart segmentation coronal results of Case 3 (female). Sub-image A shows a representative coronal slice of the outlier segmentation applied to the 1000 s/mm^2 b--value DWI image. The normal hyperintense organs are correctly removed and no lesions are detected on the skeleton. There are small regions detected that are not part of the skeleton: spinal nerves, knees, neck glands (B) and sinus liquid (C). Sub-images D and E display the same zoomed coronal slices of B and C, using the connected-component approach applied to the 1000 s/mm^2 b-value image. The neck glands are removed but the sinus liquid remains.

On Case 4, only E3 and E4 segmented a very small area on the iliac bone (figure 6.24, sub-images A and B). As seen on male's Case 8, this lesion is very typical when medulla biopsy is performed and can be retained for months. In this case in particular, the biopsy was performed on the same day as the MRI acquisitions. Thus, this is not a real MM lesion, although segmented by two radiologists. The smart segmentation algorithm did not detect this area (figure 6.24, sub-image C), which increases merit and potential on this tool to differentiate between suspicious and MM lesions. This does not mean that the smart algorithm does not detect suspicious small areas on the iliac bone. For instance, on Case 7, two radiologists (E2 and E4) found a very small area on the iliac bone (figure 6.25, sub-image B and C). As displayed on A of figure 6.25 and comparing with figure 6.12, this suspicious lesion is very similar to a lesion left by the biopsy needle. The smart segmentation algorithm detected this small area using k=3 on the 1000 s/mm^2 b-value image (sub--image D). Plus, records show that this patient was not submitted to any procedure that would cause this lesion, so it is most probably a MM lesion. However, this area disappeared when incorporating T1w information. Not even using 1% percentile on the connected-component approach was possible to detect it. In fact, this area is clearly hyperintense on DWI; on the other hand, as seen on sub-image A of figure 6.25 (T1w image), the region is not darker than the muscle. Thus, considering the criteria defined to detected MM lesions, the algorithm upholds that this area is not a lesion, even though there are no records of a biopsy procedure.

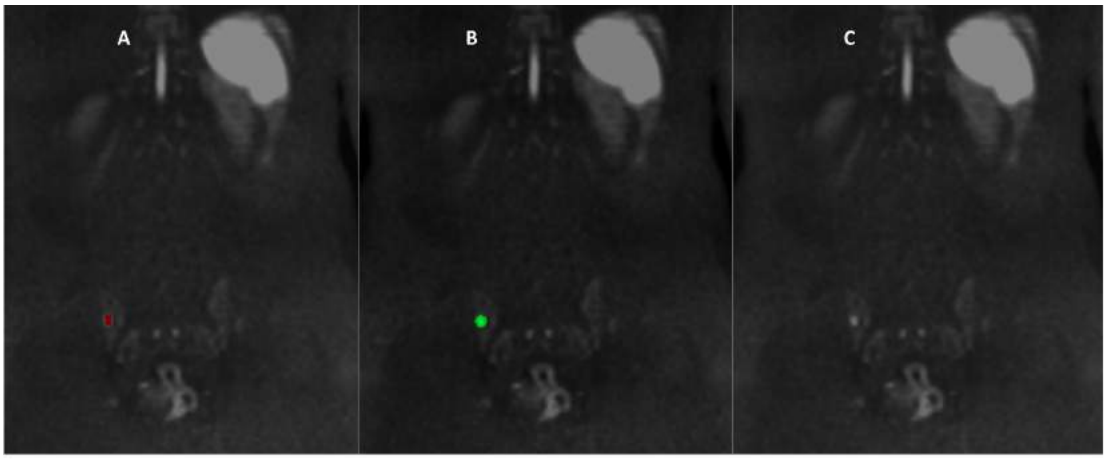
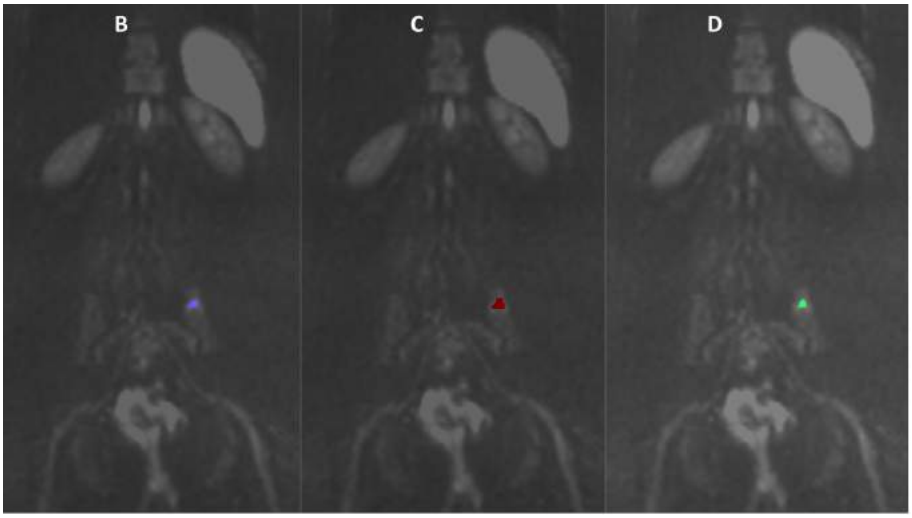


Figure 6.24: Side-by-side comparison of manual and smart segmentation of Case 4 (female). A and B show a representative zoomed coronal slice where E3 and E4 segmented a small area of the iliac bone, respectively. The smart segmentation did not detect that area by the outlier removal, using k=3 on the 1000 s/mm^2 b-value image, as seen on C. No other areas were detected by the radiologists. It is also visible that the spleen and kidneys were correctly removed by the smart algorithm.

a



b

Figure 6.25: Side-by-side comparison of manual and smart segmentation of Case 7 (female). (a) Original DWI and T1w images: A show a representative zoomed coronal slice of 1000 s/mm^2 b-value DWI and T1w images. It is visible a small hyperintense on DWI and dark on T1w area on the iliac bone. (b) Segmentation: Sub-images B, C and D show E2, E3 and smart segmentation, respectively. The smart segmentation is referred to using just the outlier segmentation, using k=3 on 1000 s/mm^2 b-value image.

Overall, both for male and female validation images, k value of 3 or higher was found as the best to include all suspicious areas detected by radiologists. We verified that the smaller the number of lesions, the higher the k value (up to 4). This is justified based on the fact that the smaller the number of lesions, the highest the probability of it being considered as an outlier. Thus, the threshold limit needs to be higher. As for the percentile used on the connected-component approach, the higher the number of reported lesions, the lower the percentile used, to guarantee that all areas would be included. On images where a small number of lesions was reported by the radiologists, percentiles of 10th to 25th were often used. However, the biggest the percentile, the more confidant it is a lesion since, in its majority, its T1w intensity is smaller than of the muscle.

6.2.3 Similarity analysis: relation between manual and smart segmentation

Tables 6.3 and 6.4 show statistics for the connected-component approach applied to the high and computed b-values as well as for the voxel-by-voxel applied to the high and computed b-values, for the images where at least one radiologist segmented a lesion (10 male and 9 female validation images). DSC, PPV, NPV and Sensitivity ± standard deviations were computed between the Smart Algorithm (SA) and radiologists' manual segmentation. The complete tables, that include this estimations for each case, are displayed in A: A.1, A.2, A.3, A.4, A.5, A.6, A.7, A.8.

Distributions of DSC values between the SA vs Experts and Experts vs Experts are also displayed on figures 6.26, 6.27, 6.28 and 6.29 for the highest and computed b-value, male and female, applying the connected-component approach.

Table 6.3: Average DSC, PPV, NPV, Sensitivity ± standard deviation computed for the connected-component (CC) and voxel-by-voxel (VbV) approaches using high and computed b-value DWI male images. DSC computed between radiologists' manual segmentation is considered the Gold Standard (GS).

Statistics	CC high b-value	CC computed b-balue	VbV high b-value	VbV computed b-value	GS
DSC	0.17±0.15	0.12±0.17	0.08±0.10	0.05±0.08	0.33±0.29
PPV	0.15±0.12	0.13±0.16	0.09±0.08	0.09±0.12	-
NPV	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	-
Sensitivity	0.40±0.19	0.24±0.23	0.15±0.15	0.09±0.10	-

Table 6.4: Average DSC, PPV, NPV, Sensitivity ± standard deviation computed for the connected-component (CC) and voxel-by-voxel (VbV) approaches using high and computed b-value DWI female images. DSC computed between radiologists' manual segmentation is considered the Gold Standard (GS).

Statistics	CC high b-value	CC computed b-balue	VbV high b-value	VbV computed b-value	GS
DSC	0.23±0.16	0.09±0.13	0.10±0.07	0.04±0.04	0.37±0.12
PPV	0.21±0.19	0.07±0.13	0.16±0.18	0.05±0.08	-
NPV	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	-
Sensitivity	0.42±0.16	0.36±0.22	0.15±0.12	0.08±0.05	-

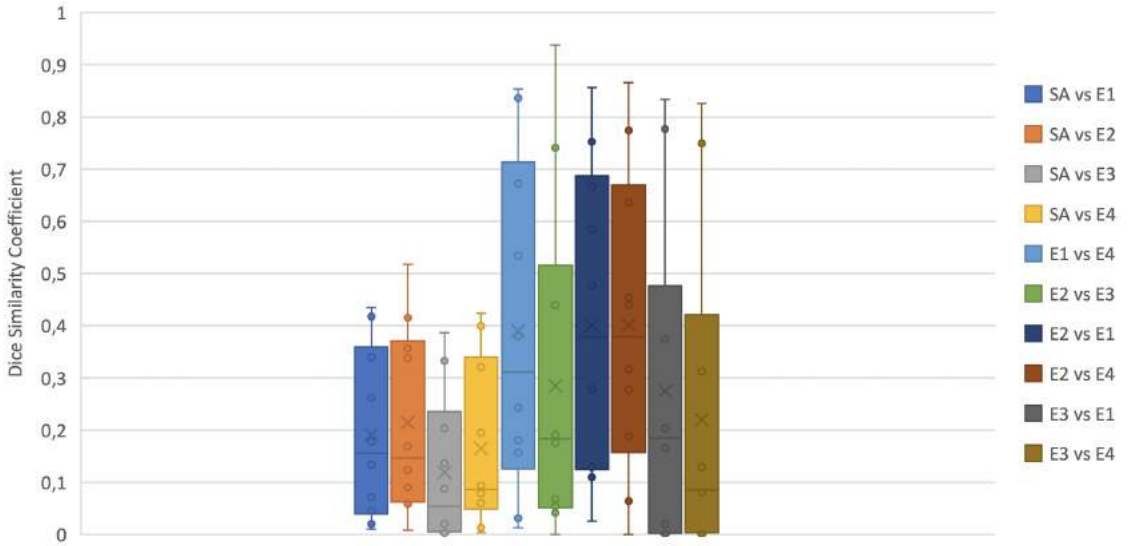


Figure 6.26: DSC distribution computed for the connected component approach using the highest b-value DWI image (800 or 1000 s/mm^2) on male images. SA vs each Expert (E1, E2, E3, E4) and Expert vs Expert values were computed. The box plot's horizontal line represents the median of the distribution, the *X* symbolizes the average, while the extremes represent the minimum and maximum values. The box plot's rectangle represents 50% of the distribution (between the first and third quartile).

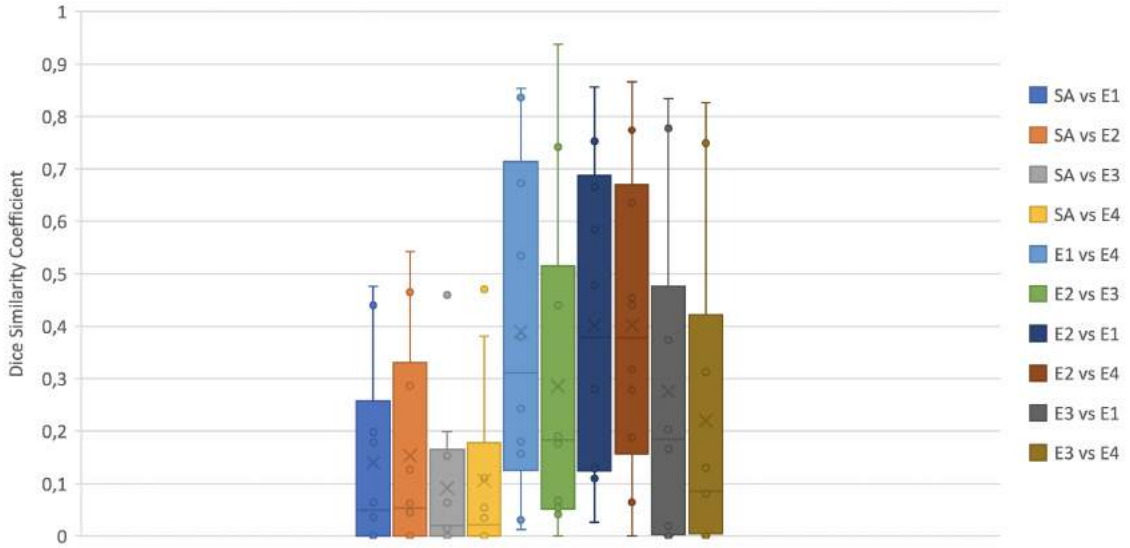


Figure 6.27: DSC distribution computed for the connected component approach using the computed b-value DWI image (1500 s/mm^2) on male images. SA vs each Expert (E1, E2, E3, E4) and Expert vs Expert values were computed. The box plot's horizontal line represents the median of the distribution, the *X* symbolizes the average, while the extremes represent the minimum and maximum values. The box plot's rectangle represents 50% of the distribution (between the first and third quartile).

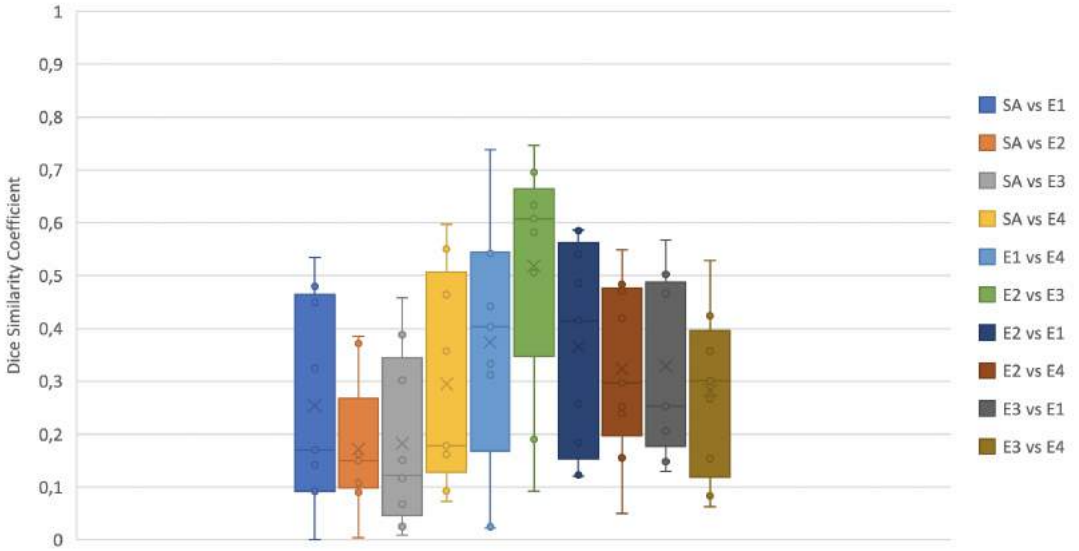


Figure 6.28: DSC distribution computed for the connected component approach using the highest b-value DWI image (800 or 1000 s/mm^2) on female images. SA vs each Expert (E1, E2, E3, E4) and Expert vs Expert values were computed. The box plot's horizontal line represents the median of the distribution, the *X* symbolizes the average, while the extremes represent the minimum and maximum values. The box plot's rectangle represents 50% of the distribution (between the first and third quartile).

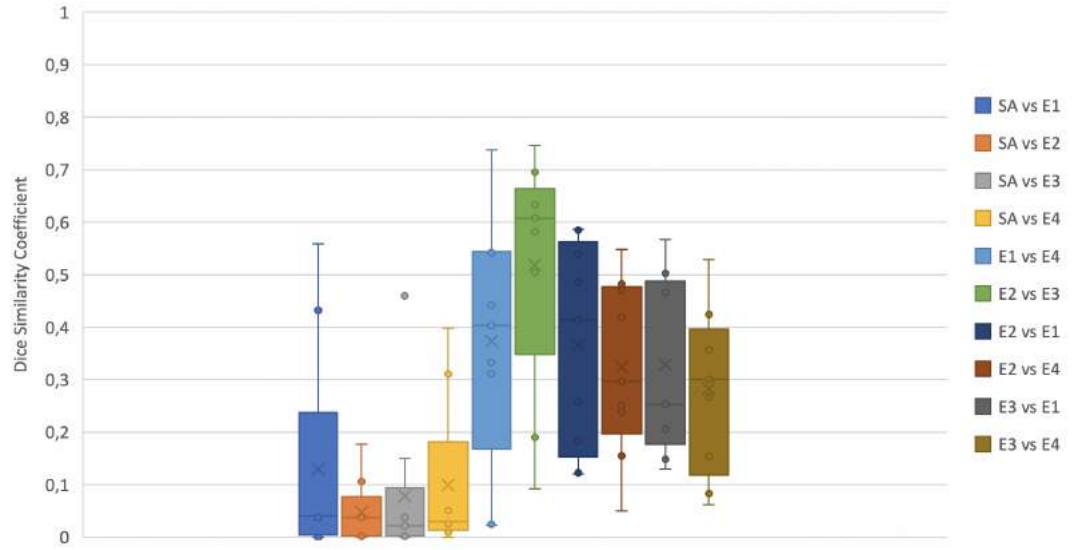


Figure 6.29: DSC distribution computed for the connected component approach using the computed b-value DWI image (1500 s/mm^2) on female images. SA vs each Expert (E1, E2, E3, E4) and Expert vs Expert values were computed. The box plot's horizontal line represents the median of the distribution, the *X* symbolizes the average, while the extremes represent the minimum and maximum values. The box plot's rectangle represents 50% of the distribution (between the first and third quartile).

Firstly, it is important to note the high disagreement between radiologists: on male validation images, in which all found lesions, 67% rate of disagreement is found; on female images, in which all found lesions, they disagree 63% of the times. The highest disagreement is found when a big number of suspicious lesions are detected, not consistent with all the radiologists (e.g. Case 5- Male; Case 16- Female, 124- Female) and when there is question whether it is a lesion or bone marrow reconversion (e.g. Case 12- Male, Case 16 - Male).

Secondly, we must remind that this similarity results do not fully express the degree of overlapping between the registered label (that contains the normal hyperintense organs) and the organs of the DWI image. However, this overlapping was visualized and measured by DSC, HD and AHD on section 6.2.2, where all the normal hyperintense organs were almost all completely removed.

Keeping in mind that the registration algorithm and removal of the normal hyperintense organs was performed for 21 male and 17 female images, being able to effectively remove those organs, we can argue that the algorithm is reproducible. Although using different body atlas (for male and female) the algorithm to build them was similar. We consider that the number of validation images is large enough to accentuate common anatomical differences and display intersubject variability, showing that the algorithm is optimized for large datasets.

Regarding the different approaches that were tested to find lesion on DWI validation images, as was expected based on the images from the previous section, the connected-component approach is the one that shows greater similarity to the gold standard, both for male and female images. The voxel-by-voxel approach applied to the computed b-value was the least succeeded. Comparing using directly acquired and computed b-values on the connected-component approach, the first was the one that presented better similarities: mean DSC 0.17±0.15, mean PPV 0.15±0.12, mean NPV 1.00±0.00, mean sensitivity 0.40±0.19 for male; mean DSC 0.23±0.16, mean PPV 0.21±0.19, mean NPV 1.00±0.00, mean sensitivity 0.42±0.16 for female.

Although it is only slightly more than half of the DSC of the manual segmentation, it is important to note that the smart segmentation algorithm is not yet fully optimized to be fairly comparable to the manual segmentation. Plus, radiologists had access to T1w, T2w, STIR and DWI images to perform segmentation, while the developed algorithm only used as input DWI and T1w images. There were intrinsic factors that were solved, as the field inhomogeneities on T1w, but others were not overcame by the algorithm. For instance, stations that are wrongly hyperintense on DWI images were detected as outliers on the Final Output DWI segmentation. In some cases, this could not be fully removed by the T1w information, since some of the voxels can lay above areas whose intensity is lower than the muscle. Also, the use of surface coils on the arms during acquisition results on smooth hyperintense on the DWI image, which is often detected as an outlier. Additionally, when the gall bladder is full of liquid (which is rather unpredictable when acquiring the image, so was not included on the normal hyperintense organs) it is also

detected as an outlier on DWI. The fact that female average DSC is greater than males may be due to the smaller number of female validation images (17 to that of 21 male).

Figure 6.26 and 6.28 are referred to using the connected-component approach to the highest b-value DWI image (800 or 1000 s/mm^2) for male and female validation, respectively. Figure 6.26 shows that experts 3 and 4 are the ones that disagree more. The algorithm shows greater agreements with experts 1 and 2, which is also coherent with the high agreement between manual segmentation of those experts. The lowest medians are found when calculating the similarity between expert 3 and any other experts. This means that this radiologist's segmentation is the one that yields more differences. The median of the DSC between the smart algorithm and expert 3 is the lowest, which confirms this difference. As for applying to the female validation images (figure 6.28), experts 1 and 3 are the ones that show the lowest median. Although experts 2 and 3 show the highest median, they also show the most scattered values, having very high DSC values (0.747) but also very low (0.092). Overall, the DSCs of the smart algorithm versus experts are also disperse, presenting DSC values between 0.597 and 0.01. The median of the DSC between the smart algorithm and expert 3 is again the lowest.

Figure 6.27 and figure 6.29 are referred to using the connected component approach using the computed b-value DWI image (1500 s/mm^2) for male and female, respectively. The mean DSC of this approach is 0.12±0.17, mean PPV 0.13±0.16, mean NPV 1.00±0.00, mean sensitivity 0.24±0.23 for male and mean DSC 0.09±0.17, mean PPV 0.13±0.16, mean NPV 1.00±0.00, mean sensitivity 0.36±0.22 for female. The smart algorithm's performance is decreased when using the computed b-value image, comparing to the acquired b-value. Although increasing the differentiability between normal and diseased tissues, for instance decreasing the intensity of the kidneys, it also increases the contribution of noise. As was seen, some smooth hyperintense areas due to the use of surface coils verified on the highest b-value image were expanded and exaggerated on the computed b-value image. This component increases the disagreement between the manual segmentation, since it introduces extra elements that were not detected by the radiologists. Although can be easily discarded as lesion, it is noisy and may induce misinterpretations.

The voxel-by-voxel approach's performance was the poorest. This is suggestive of the fact that radiologists define lesions as a whole and not as single entities. Although appearing hyperintense on DWI, lesions are heterogenous, so can also have some areas with different intensities on T1w. This was decisive on including or excluding a suspicious area from the segmentation. In fact, most of the times, the voxels' intensity was not darker than that of the Psoas muscle, being discarded as lesion.

The smart algorithm using the connected-component approach on the highest b-value DWI image seems to point out to better agreements with an overall highest mean DSC, when compared to the computed b-value DWI image. To note that this is just a preliminary result, that still lacks improvement. This further supports the notion that optimized semi-automatic registration methods combined with algorithms that quantitatively analyze DWI images can be used to assist radiologists while defining tumor burden.

Chapter 7

Conclusions and Future Work

The challenge of defining and segmenting lesions in DWI images in a semi-automatic way has been motivating the development of algorithms that have low computational time, require few user interventions and facilitate the staging process, prognosis and analysis of response. Particularly for MM, radiologists have trouble on distinguish lesions from marrow reconversion and on segmenting images that present increased number of lesions. Plus, their reproducibility is rather low. In this work, we proposed a novel smart algorithm that correctly removes the normal hyperintense organs with the use of a WB atlas and segments suspicious areas on DWI images of MM patients, incorporating T1w image information. The novelty factor in this work was the use of a WB atlas combined with an outlier detector algorithm, using as input DWI images and T1w.

Firstly, an atlas (template image) was built to represent the available population, both for male and female, based on an optimized registration algorithm that included successive rigid and deformable transformations. The normal hyperintense organs were labelled by a set of radiologists.

Secondly, an outlier detector algorithm was applied to DWI images, where normal hyperintense organs and suspicious areas were segmented. Then, the atlas was registered to that image, using the same transformation, and the label was used to remove the organs from the image. Finally, the remaining segmented areas on DWI images were overlapped with the correspondent T1w image, using an optimized rigid registration, to decide if those areas can be considered as lesions. Considering MM lesion detection criteria, two approaches were tested to support the decision: voxel-by-voxel, where each voxel's intensity was compared with the mean of Psoas' muscle, previously segmented; connected-component, where if a voxel was connected with other voxel by a single point, they would be clustered and analyzed in the same group. The algorithm was performed on directly acquired b-value DWI images (800 or 1000 s/mm^2) and computed ones (1500

s/mm^2), following literature guidance to use computed DWI to increase differentiation between tissues.

The AB performance was assessed by calculating mean DSC, PPV, NPV and sensitivity between the algorithm and radiologists' manual segmentation of normal hyperintense organs. The AB algorithm showed that the removal of the normal hyperintense organs was well succeeded, with better marks on the brain (DSC male images: 0.80±0.07; DSC female images: 0.77±0.09) and kidneys (left kidney DSC male images: 0.50±0.14; left kidney DSC female images: 0.47±0.15). The spleen and testis were the organs that were not completely removed, due to their varying sizes and mobility, respectively. Nonetheless, the spleen shows relatively good alignments to that of radiologist manual segmentation (DSC male images: 0.44±0.16; DSC female images: 0.53±0.09). As for the complete label, that joins all the hyperintense organs, DSC was 0.63±0.03 for male and 0.58±0.05 for female, when comparing the radiologist and AB segmentation. This results were very satisfactory and confirm the reliability of using AB registration algorithm as a tool to map the normal position of organs, and finally to remove them. This was not an expensive computational algorithm and did not require a numerous set of images to be trained, as powerful machine-learning algorithms would. Plus, the fact that this AB algorithm laid on image alignment and, thus, it was very intuitive, made it possible for it to be implemented from scratch during this thesis project.

The SA's performance was assessed by calculating DSC and comparing to manual segmentation of four expert radiologists. The high disagreement between radiologists was confirmed by the small mean DSC and standard deviation of the manual segmentation: 0.329±0.290 and 0.366±0.125 for male and female images, respectively. Regarding the developed algorithm, the connected-component approach applied to the highest b-value DWI image (800 or 1000 s/mm^2) was the method that presented the greatest similarity to the gold standard: mean DSC 0.17±0.15, mean PPV 0.15±0.12, mean NPV 1.00±0.00, mean sensitivity 0.40±0.19 for male; mean DSC 0.23±0.16, mean PPV 0.21±0.19, mean NPV 1.00±0.00, mean sensitivity 0.42±0.16 for female. Although not improving the manual segmentation performance, it is important to note that this is just a preliminary result, since the lesion detection criteria can still be optimized. This will be discussed on section 7.2.

Overall, the objectives of this dissertation were met. It was possible to remove the normal hyperintense organs from each DWI image used to validate the atlas algorithm, which was the primary goal. In the course of this dissertation and in obtaining the results, several challenges were faced which lead to the choice of paths. The outlier criteria to set a threshold that could fit all images was one of the chosen paths. Also, initially, our intention was just to use DWI images to find lesions, even knowing a priori that this could result in a gross segmentation. T1w image incorporation was an extra step that was take in order to make the segmentation more robust.

We believe this investigation is the first step toward developing automatic detection and segmentation algorithms for lesions, with increased sensitivity and specificity. The

time spent by radiologists while reading exams will decrease and their accuracy could, ultimately, be improved. As a final conclusion, this smart registration and segmentation algorithm has potential to assist radiologists while defining and segmenting lesions in DWI images. We believe that this can be used not only for MM patients but can be applied to a broad spectrum of diseases, as prostate carcinoma and any other malignances whose pattern on DWI is hyperintense.

7.1 Limitations

This study faced some limitations that should be considered.

Firstly, as this was a retrospective study, images had been already obtained when this study was designed. Thus, the collected images come from different scanners (1.5 T and 3 T). To our knowledge, this had no direct influence on the results, since only one 3 T image was used for the male and female atlas.

Secondly, the thickness used in these images is not the ideal one to quantify tumor burden. Instead, these sequences should employ much thinner slices, since might be discarding lesion area.

Thirdly, as was already discussed, some selected images suffered from wrongly hyperintense stations that contaminated the results. Although this is an obstacle when designing and implementing semi-automatic algorithms, the algorithm should be strong enough to interpret it, since these obstacles are very common. There is a need for the image acquisition technique to be meticulous, thus minimizing this effects. Nonetheless, this is a limitation of this dissertation, since the algorithm was not optimized to overcome these hindrances.

This study would also benefit from an increased sample in order to have a statistical meaning and also to implement an appropriated statistical test, to validate the algorithm and generalize the results.

7.2 Future Perspectives

For future perspectives, there are several steps that can be conducted:

- Restrict lesion area to the skeleton. By including bone segmentation in the atlas, this label could be used as the normal hyperintense organs' label was. Then, this would act as a narrower: if an hyperintense area on DWI is present in the skeleton registered area, and its intensity percentile is darker than that of the medium Psoas' muscle, it is considered a lesion. This would not only discard the arms' hyperintensities, but also stations that present false intensity values;

- Incorporate other sequences to increase the reliability of the segmentation. For instance, short TI inversion-recovery, opposed-phase, post-contrast T1w images;

- Compare with ADC values. After improving the smart algorithm's reliability, segmented areas could be compared with ADC maps. Each lesion area would be analyzed as a unique entity, and ADC values would be extracted. This information could be used to predict treatment outcomes or to distinguish between different MM patterns, for instance diffuse from normal;

- Develop a friendly interface to be used in the common clinical practice by radiologists. While analyzing the images, if the radiologist found lesions by visual inspection, those sequences would be imported to the interface. Several tools would be available to assist segmentation (which were already developed by code), for instance: a sliding bar to change the k value and the percentile. An eraser and marker would also be available if the radiologist wants to discard or include a specific region.

Bibliography

[1] G. P. Schmidt, M. F. Reiser, and A. Baur-Melnyk. "Whole-body MRI for the staging and follow-up of patients with metastasis." In: *European Journal of Radiology* 70.3 (2009), pp. 393–400. ISSN: 0720048X. DOI: 10.1016/j.ejrad.2009.03.045.

[2] I. Lavdas, B. Glocker, K. Kamnitsas, D. Rueckert, H. Mair, A. Sandhu, S. A. Taylor, E. O. Aboagye, and A. G. Rockall. "Fully automatic, multiorgan segmentation in normal whole body magnetic resonance imaging (MRI), using classification forests (CFs), convolutional neural networks (CNNs), and a multi-atlas (MA) approach." In: *Medical Physics* 44.10 (2017), pp. 5210–5220. ISSN: 00942405. DOI: 10.1002/mp.12492. URL: http://doi.wiley.com/10.1002/mp.12492.

[3] D. L. Pham, C. Xu, and J. L. Prince. "Current methods in medical image segmentation." In: *Annual review of biomedical engineering* 2 (2000), pp. 315–37. ISSN: 1523-9829. DOI: 10.1146/annurev.bioeng.2.1.315. URL: http://www.ncbi.nlm.nih.gov/pubmed/11701515.

[4] F. P. Oliveira and J. M. R. Tavares. *Medical image registration: A review*. 2014. DOI: 10.1080/10255842.2012.670855. URL: http://dx.doi.org/10.1080/10255842.2012.670855.

[5] L. Chen, P. Bentley, and D. Rueckert. "Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks." In: *NeuroImage: Clinical* 15 (2017), pp. 633–643. DOI: 10.1016/j.nicl.2017.06.016.

[6] Y.-H. Mah, R. Jager, C. Kennard, M. Husain, and P. Nachev. "A new method for automated high-dimensional lesion segmentation evaluated in vascular injury and applied to the human occipital lobe." In: *Cortex* 56 (2014), pp. 51–63. DOI: 10.1016/j.cortex.2012.12.008.

[7] M. D. Blackledge, D. J. Collins, N. Tunariu, M. R. Orton, A. R. Padhani, M. O. Leach, and D. M. Koh. "Assessment of treatment response by total tumor volume and global apparent diffusion coefficient using diffusion-weighted MRI in patients with metastatic bone disease: A feasibility study." In: *PLoS ONE* 9.4 (2014), pp. 1–9. ISSN: 19326203. DOI: 10.1371/journal.pone.0091779.

[8] W. M. Wells. "Medical Image Analysis – past, present, and future." In: *Medical Image Analysis* 33 (2016), pp. 1339–1351. ISSN: 13618423. DOI: 10.1016/j.media.2016.06.013.

[9] Z. Ma, J. M. R. Tavares, R. N. Jorge, and T. Mascarenhas. "A review of algorithms for medical image segmentation and their applications to the female pelvic cavity." In: *Computer Methods in Biomechanics and Biomedical Engineering* 13.2 (2010), pp. 235–246. ISSN: 10255842. DOI: 10.1080/10255840903131878.

[10] A. Mahbod, M. Chowdhury, Ö. Smedby, and C. Wang. "Automatic brain segmentation using artificial neural networks with shape context." In: *Pattern Recognition Letters* 101 (2018), pp. 74–79. ISSN: 01678655. DOI: 10.1016/j.patrec.2017.11.016. URL: https://doi.org/10.1016/j.patrec.2017.11.016.

[11] X. Zhang, W. Dou, M. Zhang, and H. Chen. "A framework of automatic brain tumor segmentation method based on information fusion of structural and functional MRI signals." In: *2016 8th IEEE International Conference on Communication Software and Networks (ICCSN)* (2016), pp. 625–629. DOI: 10.1109/ICCSN.2016.7586598. URL: http://ieeexplore.ieee.org/document/7586598/.

[12] A. Mayer, G. Zimmerman-Moreno, R. Shadmi, A. Batikoff, and H. Greenspan. "A supervised framework for the registration and segmentation of white matter fiber tracts." In: *IEEE Transactions on Medical Imaging* 30.1 (2011), pp. 131–145. ISSN: 02780062. DOI: 10.1109/TMI.2010.2067222.

[13] C. N. Devi, A. Chandrasekharan, S. V.K., and Z. C. Alex. "Automatic segmentation of infant brain MR images: With special reference to myelinated white matter." In: *Biocybernetics and Biomedical Engineering* 37.1 (2017), pp. 143–158. ISSN: 02085216. DOI: 10.1016/j.bbe.2016.11.004. URL: http://dx.doi.org/10.1016/j.bbe.2016.11.004http://linkinghub.elsevier.com/retrieve/pii/S0208521616301188.

[14] L. Tang, G. Hamarneh, and A. Celler. "Co-registration of Bone CT and SPECT Images Using Mutual Information." In: *2006 IEEE International Symposium on Signal Processing and Information Technology* (2006), pp. 116–121. DOI: 10.1109/ISSPIT.2006.270781. URL: http://ieeexplore.ieee.org/document/4042223/.

[15] F. P. Oliveira, D. B. Faria, and J. M. R. Tavares. "Automated segmentation of the incus and malleus ossicles in conventional tri-dimensional computed tomography images." In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 228.8 (2014), pp. 810–818. ISSN: 20413033. DOI: 10.1177/0954411914546123.

[16] K. A. Powell, T. Liang, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet. "Atlas-Based Segmentation of Temporal Bone Anatomy." In: *International journal of computer assisted radiology and surgery* 12.11 (2017), pp. 1937–1944. DOI: 10.1007/s11548-017-1658-6.

[17] T. Kurzendorfer, C. Forman, M. Schmidt, C. Tillmanns, A. Maier, and A. Brost. "Fully automatic segmentation of left ventricular anatomy in 3-D LGE-MRI." In: *Computerized Medical Imaging and Graphics* 59 (2017), pp. 13–27. ISSN: 18790771. DOI: 10.1016/j.compmedimag.2017.05.001.

[18] J. Shen, T. Baum, C. Cordes, B. Ott, T. Skurk, H. Kooijman, E. J. Rummeny, H. Hauner, B. H. Menze, and D. C. Karampinos. "Automatic segmentation of abdominal organs and adipose tissue compartments in water-fat MRI: Application to weight-loss in obesity." In: *European Journal of Radiology* 85.9 (2016), pp. 1613–1621. ISSN: 18727727. DOI: 10.1016/j.ejrad.2016.06.006. URL: http://dx.doi.org/10.1016/j.ejrad.2016.06.006.

[19] H. Takizawa, T. Suzuki, H. Kudo, and T. Okada. "Interactive Segmentation of Pancreases in Abdominal Computed Tomography Images and Its Evaluation Based on Segmentation Accuracy and Interaction Costs." In: *BioMed Research International* 2017 (2017), pp. 1–8. ISSN: 2314-6133. DOI: 10.1155/2017/5094592. URL: https://www.hindawi.com/journals/bmri/2017/5094592/.

[20] B. Zhou and L. Chen. "Atlas-based semi-automatic kidney tumor detection and segmentation in CT images." In: *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on*. IEEE. 2016, pp. 1397–1401. DOI: 10.1109/CISP-BMEI.2016.7852935.

[21] B. Zhou and L. Chen. "Atlas-based Semi-automatic Kidney Tumor Detection and Segmentation in CT Images." In: (2016), pp. 1397–1401.

[22] K. A. Powell, T. Liang, B. Hittle, D. Stredney, T. Kerwin, and G. J. Wiet. "Atlas-Based Segmentation of Temporal Bone Anatomy." In: *International Journal of Computer Assisted Radiology and Surgery* 12.11 (2017), pp. 1937–1944. ISSN: 18616429. DOI: 10.1007/s11548-017-1658-6.

[23] N. T. Doan, J. Orban de Xivry, and B. Macq. "Effect of inter-subject variation on the accuracy of atlas-based segmentation applied to human brain structures." In: *SPIE Medical Imaging* 7623 (2010), 76231S–76231S–10. ISSN: 16057422. DOI: 10.1117/12.845586. URL: http://link.aip.org/link/PSISDG/v7623/i1/p76231S/s1{\&}Agg=doi.

[24] A. Karlsson, J. Rosander, T. Romu, J. Tallberg, A. Grönqvist, M. Borga, and O. Dahlqvist Leinhard. "Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI." In: *Journal of Magnetic Resonance Imaging* 41.6 (2015), pp. 1558–1569. ISSN: 15222586. DOI: 10.1002/jmri.24726.

[25] M. A. Dimopoulos, J. Hillengass, S. Usmani, E. Zamagni, S. Lentzsch, F. E. Davies, N. Raje, O. Sezer, S. Zweegman, J. Shah, et al. "Role of magnetic resonance imaging in the management of patients with multiple myeloma: a consensus statement." In: *J Clin Oncol* 33.6 (2015), pp. 657–664. DOI: 10.1200/JCO.2014.57.9961.

[26] E. Terpos, M. A. Dimopoulos, and L. A. Moulopoulos. "The Role of Imaging in the Treatment of Patients With Multiple Myeloma in 2016." In: *American Society of Clinical Oncology educational book / ASCO. American Society of Clinical Oncology. Meeting* 35 (2016), e407–17. ISSN: 1548-8756. DOI: 10.14694/EDBK_159074. URL: http://www.ncbi.nlm.nih.gov/pubmed/27249748.

[27] L. A. Moulopoulos, D. Gika, A. Anagnostopoulos, K. Delasalle, D. Weber, R. Alexanian, and M. A. Dimopoulos. "Prognostic significance of magnetic resonance imaging of bone marrow in previously untreated patients with multiple myeloma." In: *Annals of Oncology* 16.11 (2005), pp. 1824–1828. DOI: 10.1093/annonc/mdi362.

[28] L. Moulopoulos, M. Dimopoulos, D Christoulas, E Kastritis, D Anagnostou, A Koureas, M Roussou, M Gavriatopoulou, M Migkou, M Iakovaki, et al. "Diffuse MRI marrow pattern correlates with increased angiogenesis, advanced disease features and poor prognosis in newly diagnosed myeloma treated with novel agents." In: *Leukemia* 24.6 (2010), p. 1206. DOI: 10.1038/leu.2010.70.

[29] V. Koutoulidis, S. Fontara, E. Terpos, F. Zagouri, D. Matsaridis, D. Christoulas, E. Panourgias, E. Kastritis, M. A. Dimopoulos, and L. A. Moulopoulos. "Quantitative Diffusion-weighted Imaging of the Bone Marrow: An Adjunct Tool for the Diagnosis of a Diffuse MR Imaging Pattern in Patients with Multiple Myeloma." In: *Radiology* 000.0 (2016), p. 160363. ISSN: 1527-1315 (Electronic). DOI: 10.1148/radiol.2016160363.

[30] A. P. Brady. "Error and discrepancy in radiology: inevitable or avoidable?" In: *Insights into imaging* 8.1 (2017), pp. 171–182. DOI: 10.1007/s13244-016-0534-1.

[31] F. E.-Z. A. El-Gamal, M. Elmogy, and A. Atwan. "Current trends in medical image registration and fusion." In: *Egyptian Informatics Journal* 17.1 (2016), pp. 99–124. ISSN: 11108665. DOI: 10.1016/j.eij.2015.09.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S111086651500047X.

[32] A. P. James and B. V. Dasarathy. "Medical image fusion: A survey of the state of the art." In: *Information Fusion* 19.1 (2014), pp. 4–19. ISSN: 15662535. DOI: 10.1016/j.inffus.2013.12.002. arXiv: 1401.0166.

[33] *Introduction to SimpleITKv4 Registration*. visited on 2018-05-10. URL: http://insightsoftwareconsortium.github.io/SimpleITK-Notebooks/Python_html/60_Registration_Introduction.html.

[34] B. Zitová and J. Flusser. "Image registration methods: A survey." In: *Image and Vision Computing* 21.11 (2003), pp. 977–1000. ISSN: 02628856. DOI: 10.1016/S0262-8856(03)00137-9.

[35] D. Rueckert and J. A. Schnabel. *Biomedical Image Processing*. Ed. by T. M. Deserno. Biological and Medical Physics, Biomedical Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 131–154. ISBN: 978-3-642-15815-5. DOI: 10.1007/978-3-642-15816-2. URL: http://link.springer.com/10.1007/978-3-642-15816-2.

[36] H. Livyatan, Z. Yaniv, and L. Joskowicz. "Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT." In: *IEEE Transactions on Medical Imaging* 22.11 (2003), pp. 1395–1406. ISSN: 02780062. DOI: 10.1109/TMI.2003.819288.

[37] S. Heger, F. Portheine, J. A. K. Ohnsorge, E. Schkommodau, and K. Radermacher. "User-interactive registration of bone with A-mode ultrasound." In: *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society* 24.2 (2003), pp. 85–95. ISSN: 0739-5175. DOI: 10.1136/bmj.1.1748.1436. URL: http://www.ncbi.nlm.nih.gov/pubmed/15825850.

[38] M. Auer, P. Regitnig, and G. A. Holzapfel. "An automatic nonrigid registration for stained histological sections." In: *IEEE Transactions on Image Processing* 14.4 (2005), pp. 475–486. ISSN: 10577149. DOI: 10.1109/TIP.2005.843756.

[39] L. K. Lee, S. C. Liew, and W. J. Thong. *Advanced Computer and Communication Engineering Technology*. Ed. by H. A. Sulaiman, M. A. Othman, M. F. I. Othman, Y. A. Rahim, and N. C. Pee. Vol. 315. Lecture Notes in Electrical Engineering. Cham: Springer International Publishing, 2015, pp. 1069–1080. ISBN: 978-3-319-07673-7. DOI: 10.1007/978-3-319-07674-4. URL: http://link.springer.com/10.1007/978-3-319-07674-4.

[40] E. Neri, D. Caramella, and C. Bartolozzi, eds. *Image Processing in Radiology*. Medical Radiology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-25915-2. DOI: 10.1007/978-3-540-49830-8. URL: http://search.ebscohost.com/login.aspx?direct=true{\&}db=edsebk{\&}AN=255954{\&}site=eds-livehttp://link.springer.com/10.1007/978-3-540-49830-8.

[41] N. Otsu. "A Threshold Selection Method from Gray-Level Histograms." In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. ISSN: 0018-9472. DOI: 10.1109/TSMC.1979.4310076. arXiv: arXiv:1011.1669v3. URL: http://ieeexplore.ieee.org/document/4310076/.

[42] J. W. Tukey. *Exploratory data analysis*. Vol. 2. Reading, Mass., 1977.

[43] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*. Vol. 589. John wiley & sons, 2005.

[44] V. Hodge and J. Austin. "A survey of outlier detection methodologies." In: *Artificial intelligence review* 22.2 (2004), pp. 85–126.

[45] V. Barnett and T. Lewis. *Outliers in statistical data*. Wiley, 1974.

[46] *RANGE, INTERQUARTILE RANGE AND BOX PLOT*. visited on 2018-05-20. URL: http://makemeanalyst.com/explore-your-data-range-interquartile-range-and-box-plot/.

[47] J. M. Prewitt. "Object enhancement and extraction." In: *Picture processing and Psychopictorics* 10.1 (1970), pp. 15–19.

[48] A. Norouzi, M. S. M. Rahim, A. Altameem, T. Saba, A. E. Rad, A. Rehman, and M. Uddin. "Medical Image Segmentation Methods, Algorithms, and Applications." In: *IETE Technical Review* 31.3 (2014), pp. 199–213. ISSN: 0256-4602. DOI: 10.1080/02564602.2014.906861. URL: http://www.tandfonline.com/doi/abs/10.1080/02564602.2014.906861.

[49] I. N. Bankman and S. Morcovescu. "Handbook of Medical Imaging. Processing and Analysis." In: *Medical Physics* 29.1 (2002), pp. 107–107. ISSN: 00942405. DOI: 10.1118/1.1429630. URL: http://doi.wiley.com/10.1118/1.1429630.

[50] J. E. Iglesias and M. R. Sabuncu. "Multi-atlas segmentation of biomedical images: A survey." In: *Medical Image Analysis* 24.1 (2015), pp. 205–219. ISSN: 13618423. DOI: 10.1016/j.media.2015.06.012. arXiv: 1412.3421. URL: http://dx.doi.org/10.1016/j.media.2015.06.012.

[51] Q. Liu, H. Mohy-ud Din, N. E. Boutagy, M. Jiang, S. Ren, J. C. Stendahl, A. J. Sinusas, and C. Liu. "Fully automatic multi-atlas segmentation of CTA for partial volume correction in cardiac SPECT/CT." In: *Physics in Medicine and Biology* 62.10 (2017), p. 3944. DOI: 10.1088/1361-6560/aa6520.

[52] A. M. Pouch, H. Wang, M. Takabe, B. M. Jackson, J. H. Gorman, R. C. Gorman, P. A. Yushkevich, and C. M. Sehgal. "Fully automatic segmentation of the mitral leaflets in 3D transesophageal echocardiographic images using multi-atlas joint label fusion and deformable medial modeling." In: *Medical image analysis* 18.1 (2014), pp. 118–129. DOI: 10.1016/j.media.2013.10.001.

[53] M. Sdika, A. Tonson, Y. Le Fur, P. J. Cozzone, and D. Bendahan. "Multi-atlas-based fully automatic segmentation of individual muscles in rat leg." In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.2 (2016), pp. 223–235. DOI: 10.1007/s10334-015-0511-6.

[54] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert. "Automated abdominal multi-organ segmentation with subject-specific atlas generation." In: *IEEE transactions on medical imaging* 32.9 (2013), pp. 1723–1730. DOI: 10.1109/TMI.2013.2265805.

[55] D. W. McRobbie, E. A. Moore, and M. J. Graves. *MRI from Picture to Proton*. Cambridge university press, 2017. DOI: 0.1017/9781107706958.

[56] P. Jezzard, P. M. Matthews, and S. M. Smith. *Functional MRI: an introduction to methods*. Oxford University Press, 2001. ISBN: 9780198527732.

[57] *Tissue Contrast (T1, T2, and Proton Density Weighting)*. visited on 2018-06-25. URL: https://clinicalgate.com/neuroimaging-structural-imaging-magnetic-resonance-imaging-computed-tomography/#f0045.

[58] *What is the spin echo used for in NMR?* visited on 2018-06-25. URL: http://medradresource.blogspot.com/2012/11/what-is-spin-echo-used-for-in-nmr.html.

[59] J. Juntu, J. Sijbers, D. Van Dyck, and J. Gielen. "Bias field correction for MRI images." In: *Computer Recognition Systems*. Springer, 2005, pp. 543–551. DOI: 10.1007/3-540-32390-2_64.

[60] A. Einstein. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.

[61] P. Hagmann, L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli. "Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond." In: *Radiographics* 26.suppl_1 (2006), S205–S223. DOI: 10.1148/rg.26si065510.

[62] E. H. de Figueiredo, A. F. Borgonovi, and T. M. Doring. "Basic concepts of MR imaging, diffusion MR imaging, and diffusion tensor imaging." In: *Magnetic Resonance Imaging Clinics* 19.1 (2011), pp. 1–22. DOI: 10.1016/j.mric.2010.10.005.

[63] J. C. Dutoit and K. L. Verstraete. "MRI in multiple myeloma: a pictorial review of diagnostic and post-treatment findings." In: *Insights into imaging* 7.4 (2016), pp. 553–569. DOI: 10.1007/s13244-016-0492-7.

[64] S. Bonekamp, C. P. Corona-Villalobos, and I. R. Kamel. "Oncologic applications of diffusion-weighted MRI in the body." In: *Journal of Magnetic Resonance Imaging* 35.2 (2012), pp. 257–279. DOI: 10.1002/jmri.22786.

[65] L. Cheng, M. D. Blackledge, D. J. Collins, M. R. Orton, N. P. Jerome, T. Feiweier, M. Rata, V. Morgan, N. Tunariu, M. O. Leach, et al. "T2-adjusted computed diffusion-weighted imaging: A novel method to enhance tumour visualisation." In: *Computers in biology and medicine* 79 (2016), pp. 92–98. DOI: 10.1016/j.compbiomed.2016.09.022.

[66] S. Gatidis, H. Schmidt, P. Martirosian, K. Nikolaou, and N. F. Schwenzer. "Apparent diffusion coefficient-dependent voxelwise computed diffusion-weighted imaging: An approach for improving SNR and reducing T2 shine-through effects." In: *Journal of Magnetic Resonance Imaging* 43.4 (2016), pp. 824–832. DOI: 10.1002/jmri.25044.

[67] M. D. Blackledge, M. O. Leach, D. J. Collins, and D.-M. Koh. "Computed diffusion-weighted MR imaging may improve tumor detection." In: *Radiology* 261.2 (2011), pp. 573–581. DOI: 10.1148/radiol.11101919.

[68] E. O. Stejskal and J. E. Tanner. "Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient." In: *The journal of chemical physics* 42.1 (1965), pp. 288–292.

[69] *How do you make a DW image?* visited on 2018-06-26. URL: http://mriquestions.com/making-a-dw-image.html.

[70] M. Poustchi-Amin, S. A. Mirowitz, J. J. Brown, R. C. McKinstry, and T. Li. "Principles and applications of echo-planar imaging: a review for the general radiologist." In: *Radiographics* 21.3 (2001), pp. 767–779.

[71] A. Padhani and A Gogbashian. "Bony metastases: assessing response to therapy with whole-body diffusion MRI." In: *Cancer imaging* 11.1A (2011), S129. DOI: 10.1102/1470-7330.2011.9034.

[72] A. Zingone and W. M. Kuehl. "Pathogenesis of monoclonal gammopathy of undetermined significance and progression to multiple myeloma." In: *Seminars in hematology*. Vol. 48. 1. Elsevier. 2011, pp. 4–12. DOI: 10.1053/j.seminhematol.2010.11.003.

[73] C Lacognata, F Crimì, A Guolo, C Varin, E De March, S Vio, A Ponzoni, G Barilà, A Lico, A Branca, et al. "Diffusion-weighted whole-body MRI for evaluation of early response in multiple myeloma." In: *Clinical radiology* 72.10 (2017), pp. 850–857. DOI: 10.1016/j.crad.2017.05.004.

[74] R. A. Kyle, D. R. Larson, T. M. Therneau, A. Dispenzieri, S. Kumar, J. R. Cerhan, and S. V. Rajkumar. "long-term Follow-up of Monoclonal Gammopathy of Undetermined Significance." In: *New England Journal of Medicine* 378.3 (2018), pp. 241–249. DOI: 10.1056/NEJMoa1709974.

[75] S. V. Rajkumar, M. A. Dimopoulos, A. Palumbo, J. Blade, G. Merlini, M.-V. Mateos, S. Kumar, J. Hillengass, E. Kastritis, P. Richardson, et al. "International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma." In: *The lancet oncology* 15.12 (2014), e538–e548. DOI: 10.1016/S1470-2045(14)70442-5.

[76] I. M. Ghobrial and O. Landgren. "How I treat smoldering multiple myeloma." In: *Blood* (2014), blood–2014.

[77] J. Caers, L. Garderet, K. M. Kortüm, M. E. O'Dwyer, N. W. van de Donk, M. Binder, S. M. Dold, F. Gay, J. Corre, Y. Beguin, et al. "An European Myeloma Network recommendation on tools for diagnosis and monitoring of multiple myeloma: what to use and when." In: *Haematologica* (2018), haematol–2018. DOI: 10.3324/haematol.2018.189159.

[78] E Terpos, M. Dimopoulos, and L. Moulopoulos. "The Role of Imaging in the Treatment of Patients With Multiple Myeloma in 2016." In: *American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Meeting*. Vol. 35. 2016, e407–17. DOI: 10.14694/EDBK_159074.

[79] E. Terpos, M. Kleber, M. Engelhardt, S. Zweegman, F. Gay, E. Kastritis, N. W. van de Donk, B. Bruno, O. Sezer, A. Broijl, et al. "European Myeloma Network guidelines for the management of multiple myeloma-related complications." In: *haematologica* 100.10 (2015), pp. 1254–1266. DOI: 10.3324/haematol.2014.117176.

[80] M. Cavo, E. Terpos, C. Nanni, P. Moreau, S. Lentzsch, S. Zweegman, J. Hillengass, M. Engelhardt, S. Z. Usmani, D. H. Vesole, et al. "Role of 18F-FDG PET/CT in the diagnosis and management of multiple myeloma and other plasma cell disorders: a consensus statement by the International Myeloma Working Group." In: *The Lancet Oncology* 18.4 (2017), e206–e217. DOI: 10.1016/S1470-2045(17)30189-4.

[81] E. Terpos, M. Kleber, M. Engelhardt, S. Zweegman, F. Gay, E. Kastritis, N. W. van de Donk, B. Bruno, O. Sezer, A. Broij, S. Bringhen, M. Beksac, A. Larocca, R. Hajek, P. Musto, H. E. Johnsen, F. Morabito, H. Ludwig, M. Cavo, H. Einsele, P. Sonneveld, M. A. Dimopoulos, and A. Palumbo. "European myeloma network guidelines for the management of multiple myeloma-related complications." In: *Haematologica* 100.10 (2015), pp. 1254–1266. ISSN: 15928721. DOI: 10.3324/haematol.2014.117176.

[82] L. Moulopoulos, T. Maris, N Papanikolaou, G Panagi, L Vlahos, and M. Dimopoulos. "Detection of malignant bone marrow involvement with dynamic contrast-enhanced magnetic resonance imaging." In: *Annals of oncology* 14.1 (2003), pp. 152–158.

[83] C Pawlyn, L Fowkes, S Otero, J. Jones, K. Boyd, F. Davies, G. Morgan, D. Collins, B Sharma, A Riddell, et al. "Whole-body diffusion-weighted MRI: a new gold standard for assessing disease burden in patients with multiple myeloma?" In: *Leukemia* 30.6 (2016), p. 1446. DOI: 10.1038/leu.2015.338.

[84] S. L. Giles, C. Messiou, D. J. Collins, V. A. Morgan, C. J. Simpkin, S. West, F. E. Davies, G. J. Morgan, and N. M. deSouza. "Whole-body diffusion-weighted MR imaging for assessment of treatment response in myeloma." In: *Radiology* 271.3 (2014), pp. 785–794. DOI: 10.1148/radiol.13131529.

[85] V. Koutoulidis, N. Papanikolaou, and L. A. Moulopoulos. "Functional and molecular MRI of the bone marrow in multiple myeloma." In: *The British journal of radiology* 91.xxxx (2018), p. 20170389.

[86] M. Horger, K. Weisel, W. Horger, A. Mroue, M. Fenchel, and M. Lichy. "Whole-body diffusion-weighted MRI with apparent diffusion coefficient mapping for early response monitoring in multiple myeloma: preliminary results." In: *American Journal of Roentgenology* 196.6 (2011), W790–W795. DOI: 10.2214/AJR.10.5979.

[87] P. A. Bonaffini, D. Ippolito, A. Casiraghi, V. Besostri, C. T. Franzesi, and S. Sironi. "Apparent diffusion coefficient maps integrated in whole-body MRI examination for the evaluation of tumor response to chemotherapy in patients with multiple

myeloma." In: *Academic radiology* 22.9 (2015), pp. 1163–1171. DOI: 10.1016/j.acra.2015.05.011.

[88] K. W. Carroll, J. F. Feller, and P. F. Tirman. "Useful internal standards for distinguishing infiltrative marrow pathology from hematopoietic marrow at MRI." In: *Journal of Magnetic Resonance Imaging* 7.2 (1997), pp. 394–398. DOI: 10.1002/jmri.1880070224.

[89] A. Małkiewicz and M. Dziedzic. "Bone marrow reconversion–imaging of physiological changes in bone marrow." In: *Polish journal of radiology* 77.4 (2012), p. 45.

[90] D. C. Zajick Jr, W. B. Morrison, M. E. Schweitzer, J. A. Parellada, and J. A. Carrino. "Benign and malignant processes: normal values and differentiation with chemical shift MR imaging in vertebral marrow." In: *Radiology* 237.2 (2005), pp. 590–596. DOI: 10.1148/radiol.2372040990.

[91] A Stäbler, A Baur, R Bartl, R Munker, R Lamerz, and M. Reiser. "Contrast enhancement and quantitative signal analysis in MR imaging of multiple myeloma: assessment of focal and diffuse growth patterns in marrow correlated with biopsies and survival rates." In: *AJR. American journal of roentgenology* 167.4 (1996), pp. 1029–1036. DOI: 10.2214/ajr.167.4.8819407.

[92] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare. "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research." In: *Journal of Digital Imaging* 31.3 (2018), pp. 290–303. ISSN: 1618-727X. DOI: 10.1007/s10278-017-0037-8. URL: https://doi.org/10.1007/s10278-017-0037-8.

[93] K. Marstal, F. Berendsen, M. Staring, and S. Klein. "SimpleElastix: A user-friendly, multi-lingual library for medical image registration." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, pp. 134–142.

[94] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig. "User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability." In: *Neuroimage* 31.3 (2006), pp. 1116–1128.

[95] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, et al. "3D Slicer as an image computing platform for the Quantitative Imaging Network." In: *Magnetic resonance imaging* 30.9 (2012), pp. 1323–1341. DOI: 10.1016/j.mri.2012.05.001.

[96] *Elastix Parameters*. visited on 2018-09-1. URL: http://elastix.isi.uu.nl/doxygen/parameter.html.

[97] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis. "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports." In: *Academic radiology* 11.2 (2004), pp. 178–189. DOI: 10.1016/S1076-6332(03)00671-8.

[98] M. Weinstock and I. M. Ghobrial. "Extramedullary multiple myeloma." In: *Leukemia & lymphoma* 54.6 (2013), pp. 1135–1141. DOI: 10.3109/10428194.2012.740562.

Appendix 1

Tables A.1, A.2, A.3, A.4 are referred to using the connected-component approach on the high and computed b-values DWI images, for male and female. Tables A.5, A.6, A.7, A.8 are referred to using the voxel-by-voxel approach on the low and computed b-values DWI images, for male and female. These include all the 21 male and 17 female images.

When no lesion was detected by a radiologist, the PPV between the SA and the manual segmentation is expected to be 0, since the SA will always segment regions. The NPV is expected to be close to 1. For the cases where DSC was assessed between radiologists who did not find lesions, there is no intersection between their segmentation since there was no segmentation. In this case, "no lesion"was written on the cell.

Table A.1: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the connected-component approach on the highest b-value DWI images (800 or 1000 s/mm^2) (male). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA avg	Radiologists' avg
1	DSC	0.417	0.415	0.386	0.424	0.836	0.741	0.753	0.774	0.777	0.749	0.411	0.772
	PPV	0.278	0.299	0.244	0.292								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.834	0.680	0.925	0.776								
2	DSC	0.340	0.338	0.333	0.320	0.853	0.938	0.856	0.866	0.834	0.826	0.333	0.862
	PPV	0.252	0.253	0.241	0.253								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.524	0.510	0.536	0.437								
3	DSC	0.000	0.017	0.017	0.017	0.000	0.576	0.000	0.686	0.000	0.583	0.013	0.308
	PPV	0.000	0.009	0.009	0.009								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.525	0.309	0.451								
4	DSC	0.000	0.000	0.002	0.002	0.000	0.000	no lesion	0.000	0.000	0.510	0.001	0.102
	PPV	0.000	0.000	0.001	0.001								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	no lesion	no lesion	0.288	0.307								
5	DSC	0.020	0.123	0.007	0.079	0.031	0.042	0.110	0.188	0.204	0.009	0.057	0.097
	PPV	0.010	0.074	0.003	0.061								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.333	0.369	0.800	0.111								
6	DSC	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.046	0.168	0.000								
7	DSC	0.000	0.000	0.000	0.001	0.000	no lesion	no lesion	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	0.078								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.312	0.000	0.062
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
9	DSC	0.134	0.169	0.088	0.094	0.157	0.068	0.129	0.454	0.374	0.089	0.121	0.212
	PPV	0.078	0.236	0.050	0.101								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.483	0.132	0.364	0.088								
10	DSC	0.000	0.000	0.001	0.001	0.000	0.000	no lesion	0.000	0.000	0.675	0.001	0.135
	PPV	0.000	0.000	0.001	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.184	0.120								
11	DSC	0.262	0.356	0.203	0.195	0.534	0.440	0.585	0.440	0.376	0.313	0.254	0.448
	PPV	0.183	0.243	0.143	0.144								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.463	0.669	0.350	0.304								
12	DSC	0.177	0.059	0.013	0.060	0.012	0.176	0.026	0.000	0.003	0.000	0.077	0.036
	PPV	0.787	0.033	0.007	0.033								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.100	0.260	0.321	0.335								
13	DSC	0.010	0.008	0.002	0.013	0.243	0.191	0.134	0.278	0.000	0.081	0.008	0.154
	PPV	0.006	0.004	0.001	0.007								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.058	0.306	0.742	0.188								
14	DSC	0.000	0.015	0.000	0.014	0.000	0.121	0.000	0.368	0.000	0.327	0.007	0.136
	PPV	0.000	0.007	0.000	0.008								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.359	0.000	0.104								
15	DSC	0.435	0.517	0.136	0.399	0.673	0.203	0.666	0.635	0.166	0.129	0.372	0.412
	PPV	0.411	0.428	0.074	0.447								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.462	0.655	0.830	0.361								
16	DSC	0.046	0.090	0.020	0.004	0.181	0.054	0.279	0.064	0.019	0.005	0.040	0.101
	PPV	0.025	0.056	0.080	0.002								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.288	0.231	0.012	0.077								
17	DSC	0.000	0.136	0.188	0.038	0.000	0.309	0.000	0.279	0.000	0.207	0.090	0.133
	PPV	0.000	0.094	0.111	0.021								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.241	0.613	0.190								
18	DSC	0.000	0.113	0.206	0.000	no lesion	0.076	0.000	0.000	0.000	0.000	0.080	0.015
	PPV	0.000	0.076	0.234	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.219	0.184	no lesion								
19	DSC	0.057	0.000	0.000	0.156	0.308	0.000	0.230	0.096	0.022	0.004	0.053	0.110
	PPV	0.036	0.000	0.000	0.146								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.134	0.000	0.000	0.169								
20	DSC	0.071	0.064	0.003	0.064	0.380	0.000	0.478	0.317	0.000	0.000	0.050	0.196
	PPV	0.043	0.037	0.001	0.048								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.202	0.237	0.556	0.096								
21	DSC	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000	no lesion	no lesion	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	no lesion								

Table A.2: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the connected-component approach on the highest b-value DWI images (800 or 1000 s/mm^2) (female). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA avg	Radiologists' avg
1	DSC	0.449	0.150	0.151	0.464	0.738	0.747	0.184	0.155	0.130	0.154	0.303	0.351
	PPV	0.439	0.085	0.083	0.467								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.459	0.611	0.772	0.461								
2	DSC	0.170	0.089	0.025	0.178	0.025	0.505	0.123	0.252	0.214	0.083	0.116	0.200
	PPV	0.096	0.056	0.014	0.186								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.768	0.219	0.129	0.171								
3	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	0.000								
4	DSC	0.000	0.000	0.000	0.002	0.000	0.000	no lesion	0.000	0.000	0.336	0.000	0.067
	PPV	0.000	0.000	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.009								
5	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
6	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
7	DSC	0.000	0.000	0.000	0.000	no lesion	0.819	0.000	0.000	0.000	0.000	0.000	0.164
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	no lesion								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.137	0.000	0.353	0.000	0.178	0.000	0.111
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	0.000								
9	DSC	0.141	0.163	0.122	0.162	0.547	0.696	0.585	0.470	0.466	0.357	0.147	0.520
	PPV	0.082	0.091	0.067	0.108								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.510	0.797	0.709	0.328								
10	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
11	DSC	0.023	0.000	0.000	0.022	0.531	0.632	0.000	0.000	0.000	0.000	0.011	0.194
	PPV	0.012	0.000	0.000	0.012								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.296	0.000	0.000	0.112								
12	DSC	0.534	0.371	0.302	0.597	0.403	0.582	0.258	0.483	0.206	0.368	0.451	0.383
	PPV	0.817	0.264	0.198	0.512								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.397	0.623	0.637	0.717								
14	DSC	0.324	0.385	0.388	0.357	0.442	0.616	0.586	0.548	0.502	0.529	0.364	0.537
	PPV	0.439	0.452	0.357	0.307								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.257	0.335	0.426	0.426								
15	DSC	0.092	0.113	0.117	0.073	0.333	0.633	0.485	0.297	0.473	0.301	0.099	0.420
	PPV	0.054	0.079	0.074	0.047								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.324	0.198	0.278	0.163								
13	DSC	0.479	0.163	0.458	0.550	0.542	0.092	0.120	0.239	0.567	0.424	0.413	0.331
	PPV	0.631	0.090	0.745	0.486								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.386	0.832	0.331	0.635								
16	DSC	0.001	0.004	0.008	0.092	0.022	0.190	0.540	0.050	0.148	0.062	0.026	0.169
	PPV	0.000	0.002	0.004	0.054								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.048	0.238	0.129	0.313								
17	DSC	0.092	0.108	0.068	0.178	0.312	0.607	0.414	0.419	0.253	0.267	0.112	0.379
	PPV	0.062	0.063	0.037	0.114								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.180	0.356	0.460	0.411								

Table A.3: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the connected-component approach on the computed b-value DWI images (1500 s/mm^2) (male). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA avg	Radiologists' avg
1	DSC	0.476	0.465	0.460	0.470	0.836	0.741	0.753	0.774	0.777	0.749	0.468	0.772
	PPV	0.348	0.374	0.314	0.357								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.754	0.615	0.862	0.688								
2	DSC	0.000	0.000	0.000	0.000	0.853	0.938	0.856	0.866	0.834	0.826	0.000	0.862
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.002	1.002	1.002	1.002								
	Sensitivity	0.002	0.002	0.003	0.002								
3	DSC	0.000	0.005	0.005	0.005	0.000	0.576	0.000	0.686	0.000	0.583	0.003	0.308
	PPV	0.000	0.002	0.002	0.002								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.340	0.200	0.295								
4	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.510	0.000	0.102
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.002	1.002	1.002	1.002								
	Sensitivity	no lesion	no lesion	0.000	0.000								
5	DSC	0.000	0.000	0.000	0.009	0.031	0.042	0.110	0.188	0.204	0.009	0.002	0.097
	PPV	0.000	0.000	0.000	0.005								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.000	0.000	0.040								
6	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	no lesion	0.000	0.000	0.000								
7	DSC	0.000	0.000	0.000	0.000	0.000	no lesion	no lesion	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.002	1.002	1.002	1.002								
	Sensitivity	no lesion	no lesion	no lesion	0.000								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.312	0.000	0.062
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	no lesion	no lesion	0.000	0.000								
9	DSC	0.035	0.062	0.025	0.034	0.157	0.068	0.129	0.454	0.374	0.089	0.039	0.212
	PPV	0.019	0.048	0.013	0.023								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.383	0.089	0.317	0.066								
10	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.675	0.000	0.135
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	no lesion	no lesion	0.000	0.000								
11	DSC	0.197	0.287	0.153	0.110	0.534	0.440	0.585	0.440	0.376	0.313	0.187	0.448
	PPV	0.147	0.208	0.115	0.087								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.300	0.462	0.228	0.149								
12	DSC	0.178	0.045	0.014	0.054	0.012	0.176	0.026	0.000	0.003	0.000	0.073	0.036
	PPV	0.865	0.026	0.007	0.030								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.099	0.182	0.321	0.276								
13	DSC	0.000	0.000	0.000	0.000	0.243	0.191	0.134	0.278	0.000	0.081	0.000	0.154
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.003	1.003	1.003	1.003								
	Sensitivity	0.004	0.129	0.581	0.072								
14	DSC	0.000	0.009	0.000	0.010	0.000	0.121	0.000	0.368	0.000	0.327	0.005	0.136
	PPV	0.000	0.004	0.000	0.005								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.402	0.000	0.129								
15	DSC	0.440	0.542	0.200	0.381	0.673	0.203	0.666	0.635	0.166	0.129	0.390	0.412
	PPV	0.546	0.566	0.115	0.585								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.368	0.519	0.769	0.282								
16	DSC	0.064	0.127	0.064	0.000	0.181	0.054	0.279	0.064	0.019	0.005	0.064	0.101
	PPV	0.035	0.082	0.292	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.348	0.286	0.036	0.000								
17	DSC	0.000	0.072	0.039	0.000	0.000	0.309	0.000	0.279	0.000	0.207	0.028	0.133
	PPV	0.000	0.044	0.021	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.202	0.214	0.000								
18	DSC	0.000	0.000	0.000	0.000	no lesion	0.076	0.000	0.000	0.000	0.000	0.000	0.015
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	no lesion								
19	DSC	0.000	0.000	0.000	0.006	0.308	0.000	0.230	0.096	0.022	0.004	0.001	0.110
	PPV	0.000	0.000	0.000	0.004								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.000	0.000	0.017								
20	DSC	0.009	0.006	0.000	0.000	0.380	0.000	0.478	0.317	0.000	0.000	0.004	0.196
	PPV	0.005	0.003	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.092	0.082	0.000	0.000								
21	DSC	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000	no lesion	no lesion	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	no lesion								

Table A.4: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the connected-component approach on the computed b-value DWI images (1500 s/mm^2) (female). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA avg	Radiologists' avg
1	DSC	0.009	0.003	0.003	0.009	0.738	0.747	0.184	0.155	0.130	0.154	0.006	0.351
	PPV	0.005	0.002	0.002	0.005								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.213	0.512	0.654	0.213								
2	DSC	0.042	0.012	0.003	0.018	0.025	0.505	0.123	0.252	0.214	0.083	0.019	0.200
	PPV	0.022	0.007	0.001	0.012								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.668	0.097	0.047	0.041								
3	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	0.000								
4	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.336	0.000	0.067
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
5	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
6	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.003	1.003	1.003	1.003								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
7	DSC	0.000	0.000	0.000	0.000	no lesion	0.819	0.000	0.000	0.000	0.000	0.000	0.164
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	no lesion								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.137	0.000	0.353	0.000	0.178	0.000	0.111
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	0.000								
9	DSC	0.043	0.049	0.037	0.050	0.547	0.696	0.585	0.470	0.466	0.357	0.045	0.520
	PPV	0.022	0.025	0.019	0.028								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.447	0.710	0.654	0.272								
10	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
11	DSC	0.033	0.000	0.000	0.040	0.531	0.632	0.000	0.000	0.000	0.000	0.018	0.194
	PPV	0.018	0.000	0.000	0.023								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.330	0.000	0.000	0.162								
12	DSC	0.559	0.177	0.151	0.311	0.403	0.582	0.258	0.483	0.206	0.368	0.299	0.383
	PPV	0.482	0.102	0.084	0.194								
	NPV	1.000	1.001	1.001	1.001								
	Sensitivity	0.666	0.684	0.766	0.776								
14	DSC	0.041	0.042	0.029	0.030	0.442	0.616	0.586	0.548	0.502	0.529	0.035	0.537
	PPV	0.022	0.022	0.015	0.015								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.287	0.366	0.397	0.477								
15	DSC	0.000	0.000	0.000	0.000	0.333	0.633	0.485	0.297	0.473	0.301	0.000	0.420
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.000	0.000	0.000	0.000								
13	DSC	0.432	0.107	0.460	0.398	0.542	0.092	0.120	0.239	0.567	0.424	0.349	0.331
	PPV	0.461	0.057	0.589	0.305								
	NPV	1.000	1.001	1.000	1.000								
	Sensitivity	0.407	0.761	0.377	0.575								
16	DSC	0.000	0.002	0.002	0.026	0.022	0.190	0.540	0.050	0.148	0.062	0.008	0.169
	PPV	0.000	0.001	0.001	0.014								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.060	0.238	0.064	0.146								
17	DSC	0.038	0.038	0.022	0.052	0.312	0.607	0.414	0.419	0.253	0.267	0.037	0.379
	PPV	0.021	0.020	0.011	0.028								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.225	0.416	0.522	0.380								

Table A.5: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the voxel-by-voxel approach on the lowest b-value DWI images (800 or 1000 s/mm^2) (male). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA avg	Radiologists' avg
1	DSC	0.204	0.198	0.196	0.204	0.836	0.741	0.753	0.774	0.777	0.749	0.200	0.772
	PPV	0.148	0.157	0.133	0.153								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.328	0.265	0.373	0.302								
2	DSC	0.315	0.312	0.314	0.292	0.853	0.938	0.856	0.866	0.834	0.826	0.308	0.862
	PPV	0.260	0.261	0.253	0.261								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.398	0.387	0.414	0.332								
3	DSC	0.000	0.003	0.003	0.003	0.000	0.576	0.000	0.686	0.000	0.583	0.002	0.308
	PPV	0.000	0.002	0.002	0.002								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.029	0.017	0.025								
4	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.510	0.000	0.102
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
5	DSC	0.000	0.006	0.000	0.002	0.031	0.042	0.110	0.188	0.204	0.009	0.002	0.097
	PPV	0.000	0.005	0.000	0.003								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.006	0.000	0.002								
6	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	0.000								
7	DSC	0.000	0.000	0.000	0.001	0.000	no lesion	no lesion	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	0.039								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.312	0.000	0.062
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
9	DSC	0.102	0.116	0.069	0.052	0.157	0.068	0.129	0.454	0.374	0.089	0.085	0.212
	PPV	0.066	0.243	0.043	0.079								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.230	0.076	0.175	0.039								
10	DSC	0.000	0.000	0.001	0.001	0.000	0.000	no lesion	0.000	0.000	0.675	0.000	0.135
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.032	0.021								
11	DSC	0.119	0.158	0.092	0.096	0.534	0.440	0.585	0.440	0.376	0.313	0.116	0.448
	PPV	0.071	0.093	0.055	0.059								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.361	0.514	0.272	0.252								
12	DSC	0.037	0.024	0.007	0.062	0.012	0.176	0.026	0.000	0.003	0.000	0.032	0.036
	PPV	0.582	0.018	0.004	0.043								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.019	0.036	0.045	0.111								
13	DSC	0.001	0.001	0.000	0.002	0.243	0.191	0.134	0.278	0.000	0.081	0.001	0.154
	PPV	0.001	0.000	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.006	0.024	0.097	0.027								
14	DSC	0.000	0.003	0.000	0.003	0.000	0.121	0.000	0.368	0.000	0.327	0.001	0.136
	PPV	0.000	0.002	0.000	0.002								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.027	0.000	0.008								
15	DSC	0.056	0.074	0.017	0.053	0.673	0.203	0.666	0.635	0.166	0.129	0.050	0.412
	PPV	0.079	0.086	0.010	0.092								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.044	0.065	0.055	0.037								
16	DSC	0.047	0.088	0.019	0.005	0.181	0.054	0.279	0.064	0.019	0.005	0.040	0.101
	PPV	0.026	0.056	0.086	0.002								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.266	0.205	0.011	0.077								
17	DSC	0.000	0.049	0.066	0.005	0.000	0.309	0.000	0.279	0.000	0.207	0.030	0.133
	PPV	0.000	0.042	0.044	0.003								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.059	0.134	0.016								
18	DSC	0.000	0.073	0.033	0.000	no lesion	0.076	0.000	0.000	0.000	0.000	0.026	0.015
	PPV	0.000	0.089	0.103	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.062	0.020	no lesion								
19	DSC	0.008	0.000	0.000	0.007	0.308	0.000	0.230	0.096	0.022	0.004	0.004	0.110
	PPV	0.009	0.000	0.000	0.018								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.008	0.000	0.000	0.004								
20	DSC	0.000	0.000	0.000	0.000	0.380	0.000	0.478	0.317	0.000	0.000	0.000	0.196
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.000	0.000	0.000								
21	DSC	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000	no lesion	no lesion	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	no lesion								

Table A.6: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the voxel-by-voxel approach on the lowest b-value DWI images (800 or 1000 s/mm^2) (female). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA average	Radiologists' average
1	DSC	0.198	0.034	0.033	0.205	0.738	0.747	0.184	0.155	0.130	0.154	0.117	0.351
	PPV	0.130	0.137	0.018	0.017								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.413	0.387	0.484	0.410								
2	DSC	0.064	0.037	0.036	0.054	0.025	0.505	0.123	0.252	0.214	0.083	0.048	0.200
	PPV	0.040	0.028	0.022	0.083								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.167	0.057	0.104	0.040								
3	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	0.000								
4	DSC	0.000	0.000	0.000	0.001	0.000	0.000	no lesion	0.000	0.000	0.336	0.000	0.067
	PPV	0.000	0.000	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.003								
5	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
6	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
7	DSC	0.000	0.000	0.000	0.000	no lesion	0.819	0.000	0.000	0.000	0.000	0.000	0.164
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	no lesion								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.137	0.000	0.353	0.000	0.178	0.000	0.111
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	0.000								
9	DSC	0.055	0.053	0.046	0.046	0.547	0.696	0.585	0.470	0.466	0.357	0.050	0.520
	PPV	0.053	0.044	0.035	0.067								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.057	0.067	0.065	0.035								
10	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
11	DSC	0.000	0.000	0.000	0.000	0.531	0.632	0.000	0.000	0.000	0.000	0.000	0.194
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.009	0.000	0.000	0.003								
12	DSC	0.100	0.171	0.159	0.189	0.403	0.582	0.258	0.483	0.206	0.368	0.155	0.383
	PPV	0.779	0.342	0.253	0.569								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.054	0.114	0.116	0.113								
14	DSC	0.208	0.250	0.301	0.272	0.442	0.616	0.586	0.548	0.502	0.529	0.258	0.537
	PPV	0.476	0.477	0.414	0.342								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.133	0.169	0.237	0.226								
15	DSC	0.020	0.034	0.030	0.017	0.333	0.633	0.485	0.297	0.473	0.301	0.025	0.420
	PPV	0.011	0.019	0.016	0.009								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.286	0.208	0.269	0.143								
13	DSC	0.096	0.058	0.095	0.117	0.542	0.092	0.120	0.239	0.567	0.424	0.092	0.331
	PPV	0.311	0.040	0.407	0.208								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.057	0.110	0.054	0.081								
16	DSC	0.000	0.001	0.001	0.110	0.022	0.190	0.540	0.050	0.148	0.062	0.028	0.169
	PPV	0.000	0.001	0.001	0.079								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.029	0.008	0.177								
17	DSC	0.061	0.077	0.059	0.150	0.312	0.607	0.414	0.419	0.253	0.267	0.087	0.379
	PPV	0.073	0.067	0.039	0.161								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.052	0.092	0.120	0.141								

Table A.7: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the voxel-by-voxel approach on the computed b-value DWI images (1500 s/mm^2) (male). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA avg	Radiologists' avg
1	DSC	0.284	0.261	0.301	0.276	0.836	0.741	0.753	0.774	0.777	0.749	0.280	0.772
	PPV	0.292	0.311	0.276	0.302								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.277	0.224	0.331	0.254								
2	DSC	0.000	0.000	0.000	0.000	0.853	0.938	0.856	0.866	0.834	0.826	0.000	0.862
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.001	0.001	0.001	0.001								
3	DSC	0.000	0.000	0.000	0.000	0.000	0.576	0.000	0.686	0.000	0.583	0.000	0.308
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.013	0.007	0.011								
4	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.510	0.000	0.102
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	no lesion	no lesion	0.000	0.000								
5	DSC	0.000	0.000	0.000	0.002	0.031	0.042	0.110	0.188	0.204	0.009	0.001	0.097
	PPV	0.000	0.000	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.000	0.000	0.003								
6	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	0.000								
7	DSC	0.000	0.000	0.000	0.000	0.000	no lesion	no lesion	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	no lesion	no lesion	no lesion	0.000								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.312	0.000	0.062
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
9	DSC	0.038	0.064	0.024	0.027	0.157	0.068	0.129	0.454	0.374	0.089	0.038	0.212
	PPV	0.021	0.072	0.013	0.025								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.188	0.058	0.140	0.031								
10	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.675	0.000	0.135
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
11	DSC	0.068	0.096	0.052	0.042	0.534	0.440	0.585	0.440	0.376	0.313	0.064	0.448
	PPV	0.039	0.055	0.030	0.025								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.251	0.385	0.188	0.135								
12	DSC	0.037	0.026	0.008	0.055	0.012	0.176	0.026	0.000	0.003	0.000	0.031	0.036
	PPV	0.707	0.021	0.004	0.040								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.019	0.034	0.045	0.087								
13	DSC	0.000	0.000	0.000	0.000	0.243	0.191	0.134	0.278	0.000	0.081	0.000	0.154
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.006	0.017	0.097	0.018								
14	DSC	0.000	0.002	0.000	0.002	0.000	0.121	0.000	0.368	0.000	0.327	0.001	0.136
	PPV	0.000	0.001	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.038	0.000	0.012								
15	DSC	0.061	0.087	0.046	0.051	0.673	0.203	0.666	0.635	0.166	0.129	0.061	0.412
	PPV	0.329	0.358	0.046	0.378								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.033	0.049	0.047	0.028								
16	DSC	0.058	0.127	0.056	0.000	0.181	0.054	0.279	0.064	0.019	0.005	0.060	0.101
	PPV	0.033	0.085	0.307	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.266	0.247	0.031	0.000								
17	DSC	0.000	0.040	0.011	0.000	0.000	0.309	0.000	0.279	0.000	0.207	0.013	0.133
	PPV	0.000	0.026	0.007	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.087	0.048	0.000								
18	DSC	0.000	0.020	0.017	0.000	no lesion	0.076	0.000	0.000	0.000	0.000	0.009	0.015
	PPV	0.000	0.022	0.044	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.019	0.010	no lesion								
19	DSC	0.001	0.000	0.000	0.002	0.308	0.000	0.230	0.096	0.022	0.004	0.001	0.110
	PPV	0.001	0.000	0.000	0.001								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.004	0.000	0.000	0.003								
20	DSC	0.001	0.000	0.000	0.000	0.380	0.000	0.478	0.317	0.000	0.000	0.000	0.196
	PPV	0.001	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.007	0.000	0.000	0.000								
21	DSC	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000	no lesion	no lesion	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	no lesion								

Table A.8: DSC, PPV, NPV, Sensitivity and averages (avg) computed for the SA and Experts, using the voxel-by-voxel approach on the computed b-value DWI images (1500 s/mm^2) (female). This table includes the results for the images where at least one of the radiologists did not find lesions.

Case	Statistics	SA vs E1	SA vs E2	SA vs E3	SA vs E4	E1 vs E4	E2 vs E3	E2 vs E1	E2 vs E4	E3 vs E1	E3 vs E4	SA average	Radiologists' average
1	DSC	0.020	0.023	0.024	0.020	0.738	0.747	0.184	0.155	0.130	0.154	0.021	0.351
	PPV	0.015	0.012	0.013	0.016								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.027	0.152	0.199	0.026								
2	DSC	0.016	0.009	0.005	0.022	0.025	0.505	0.123	0.252	0.214	0.083	0.013	0.200
	PPV	0.009	0.005	0.003	0.017								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.142	0.039	0.051	0.032								
3	DSC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.000
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	no lesion	0.000								
4	DSC	0.000	0.000	0.000	0.000	0.000	0.000	no lesion	0.000	0.000	0.336	0.000	0.067
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	0.000	0.000								
5	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
6	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.003	1.003	1.003	1.003								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
7	DSC	0.000	0.000	0.000	0.000	no lesion	0.819	0.000	0.000	0.000	0.000	0.000	0.164
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	no lesion								
8	DSC	0.000	0.000	0.000	0.000	0.000	0.137	0.000	0.353	0.000	0.178	0.000	0.111
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	0.000	0.000	0.000								
9	DSC	0.011	0.011	0.010	0.012	0.547	0.696	0.585	0.470	0.466	0.357	0.011	0.520
	PPV	0.006	0.006	0.005	0.008								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.049	0.070	0.071	0.030								
10	DSC	0.000	0.000	0.000	0.000	no lesion	no lesion	no lesion	no lesion	no lesion	no lesion	0.000	-
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	no lesion	no lesion	no lesion	no lesion								
11	DSC	0.000	0.000	0.000	0.000	0.531	0.632	0.000	0.000	0.000	0.000	0.000	0.194
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.000	0.000	0.000								
12	DSC	0.137	0.098	0.084	0.126	0.403	0.582	0.258	0.483	0.206	0.368	0.111	0.383
	PPV	0.290	0.081	0.062	0.134								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.090	0.122	0.128	0.120								
14	DSC	0.031	0.034	0.027	0.027	0.442	0.616	0.586	0.548	0.502	0.529	0.029	0.537
	PPV	0.018	0.019	0.014	0.014								
	NPV	1.001	1.001	1.001	1.001								
	Sensitivity	0.120	0.158	0.193	0.222								
15	DSC	0.000	0.000	0.000	0.000	0.333	0.633	0.485	0.297	0.473	0.301	0.000	0.420
	PPV	0.000	0.000	0.000	0.000								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.000	0.000	0.000								
13	DSC	0.099	0.046	0.103	0.104	0.542	0.092	0.120	0.239	0.567	0.424	0.088	0.331
	PPV	0.281	0.030	0.383	0.165								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.060	0.098	0.060	0.076								
16	DSC	0.000	0.001	0.001	0.061	0.022	0.190	0.540	0.050	0.148	0.062	0.016	0.169
	PPV	0.000	0.001	0.001	0.044								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.000	0.029	0.008	0.104								
17	DSC	0.023	0.022	0.014	0.041	0.312	0.607	0.414	0.419	0.253	0.267	0.025	0.379
	PPV	0.014	0.012	0.007	0.025								
	NPV	1.000	1.000	1.000	1.000								
	Sensitivity	0.056	0.093	0.122	0.118								

ANNEX I

Annex 1 - Ethics Committee Approval

Comissão de Ética
Prof. Doutor José Manuel Cardoso da Costa (Presidente)
Prof. Doutor Adelino Cardoso
Prof. Doutor António Jacinto
Prof. Doutor José Cunha Vaz
Profª. Doutora Leonor Parreira
Prof. Doutor Mário Miguel Rosa
Dra. Paula Martinho da Silva

Fundação Champalimaud

Lisboa, 18 de julho de 2018

Projeto: DWI Body Atlas – Prof. Cristina João "Atlas based semi-automatic segmentation of Whole-Body DWI Images: Quantification of Tumor Burden."

Após análise dos documentos que servem de base à candidatura, a Comissão de Ética aprovou condicionalmente este projeto, com a indicação de que devem ser incluídas as recomendações constantes no parecer dos Profs. Drs. Leonor Parreira e António Jacinto. As alterações requeridas deverão ser enviadas a este Conselho.

Upon review of the proposal documents, the Ethics Committee approved conditionally this project, after the project takes in account the recommendations made by Drs. Leonor Parreira and António Jacinto which should be sent to the Ethics Committee.

O Presidente da Comissão de Ética,
The President of the Ethics Committee,

Prof. Dr. José Manuel Cardoso da Costa, J.D.

Fundação D. Anna de Sommar Champalimaud e
Dr. Carlos Montez Champalimaud
Fundada por António Champalimaud

Avenida Brasília
1400-038 Lisboa, Portugal
T (+351) 210 480 200
F (+351) 210 480 299
www.fchampalimaud.org

Comissão de Ética
Prof. Doutor José Manuel Cardoso da Costa (Presidente)
Prof. Doutor Adelino Cardoso
Prof. Doutor António Jacinto
Prof. Doutor José Cunha Vaz
Profª. Doutora Leonor Parreira
Prof. Doutor Mário Miguel Rosa
Dra. Paula Martinho da Silva

Fundação Champalimaud

Projeto: DWI Body Atlas – Prof. Cristina João. "Atlas based semi-automatic segmentation of Whole-Body DWI Images: Quantification of Tumor Burden."

Investigadora Principal: Prof.ª Cristina João

Co-investigadores:

Sílvia Alexandra Dias Almeida

Prof. Nickolas Papanikolaou

Prof.ª Lisitskaya Maria

Prof. Francisco Oliveira

Dr. João Lourenço

Dr.ª Joana Ip

Dr. Aycan Uysal

Eng. João Santinha

Apreciação:

Trata-se de um estudo retrospetivo sem intervenção cujo objetivo primário é a criação de um atlas imagiológico (DWI) de corpo inteiro em doentes com cancro e um algoritmo computacional tendo em vista a correta quantificação de lesões ósseas. Este algoritmo será validado em pessoas saudáveis e doentes com cancro em diferentes estádios da doença e comparado com métodos tradicionais.

O estudo, que constituirá a tese final de mestrado integrado em Engenharia Biomédica da estudante Alexandra Almeida, envolve a análise retrospetiva de uma série de 93 doentes com diversas neoplasias, durante um período de 6 meses, não envolvendo qualquer intervenção (apenas análise de imagens previamente obtidas e anonimizadas), pelo que a investigadora principal solicita à CE dispensa de consentimento informado.

Se tiver sucesso o estudo, claramente apresentado e justificado na nova proposta enviada à CES, é seguramente pertinente e com potencial valor para futuros doentes.

Os investigadores propõem o recurso ao consentimento informado geral sobre investigação científica para doentes que o tenham assinado e, no caso de doentes mais antigos, falecidos ou de contacto impossível, solicitam que seja autorizada a dispensa total de consentimento. Para tal, as imagens serão anonimizadas.

Assim:

Tratando-se de um estudo retrospetivo de imagens (radiológicas),potencialmente útil na avaliação diagnóstica de futuros doentes, sem qualquer intervenção ou prejuízo para os doentes envolvidos, e face à potencial dificuldade de obter retrospetivamente consentimento específico, é autorizada a dispensa de consentimento informado, uma vez que foram asseguradas pelos investigadores as condições institucionais necessárias à adequada anonimização de dados.

Lisboa, 1 agosto 2018

Prof. Dra. Leonor Parreira